



Proceedings of The Fourth International Conference on Electronic Business

Xiyuan Hotel, Beijing, China
5-9 December 2004
ISSN: 1683-0040





**Proceedings of
The Fourth
International Conference on
Electronic Business**

*"Shaping Business Strategy in a
Networked World"*

Edited by

Jian Chen,
Tsinghua University, China

ICEB 2004
5-9 December 2004
Xiyuan Hotel, Beijing, China

ISSN: 1683-0040

Reference to this book should be made as follows: Chen, J. Ed. (2004). *Proceedings of the 4th International Conference on Electronic Business* (ISSN: 1683-0040). ICEB, Beijing, China, December 5-9, 1390 pages.

Table of Contents

Program	iii
A Framework to Support Coalition Formation in Supply Chain Collaboration	1
A Study of Benefits of Information Sharing between Production Systems and Service Centers ..	7
A Study of Corporate Managerial Module and Information System Orienting Agile Virtual Enterprises	14
A Study of Single-vendor and Multiple-retailers Pricing-Ordering Strategy under Group-Buying Online Auction	19
An E-Logistics Model for Effective Collaborative Commerce	24
An XCS-Based Intelligent Searching Model for Cross-Organization Identity Management in Web Service	28
Assessment of Service Quality of Logistics Center in Chain Stores Based on Fuzzy Logic and TOPSIS	34
Asymmetric Demand Information and Channel Profits When Retailers Compete	38
Bidding Strategies Analysis for Procurement Combinatorial Auctions	41
Business Interactions with Web Services	46
Collaboration Strategy of Business-to-Business Enhancing Customer Satisfaction	52
Collaborative Planning, Forecasting and Replenishment: Demand Planning in Supply Chain Management	58
Cooperation in Reverse Logistics	63
Coordination Mechanisms in the Supply Chain Based E-business	68
Decision Model for COTS Component Procurement Based on Case-based Retrieval and Goal Programming	73
Design of an E-SCM system for Auto-Parts Companies	77
Development Stages of Supply Chain Management in E-Business Environment	82
E-Booking Control Problem in Sea Cargo Logistics	87
E-Logistics Comparative Positioning Model: A Multi-National Enterprise Airline Study	93
E-supply Chain's Implementation in a Transit Economy	101
Experimental Study on Effect of Information Behavior by Media Representation Patterns of E-Commerce System	110
Generic Supply Chain Management System	114
Hierarchical Structuring of a Workflow Model in Petri-Net	117
Manufacturing Enterprises Value Chain Model Analysis within the Context of the Network Economy	122
On Inventory Strategies of Online Retailers	127
On-Demand Business Process Integration Based on Intelligent Web Services	132
A Two Phase Verification Algorithm for Cyclic Workflow Graphs	137
Pricing Policy for Supply Chain in the Electronic Commerce Environment	144
Recomposing of Industrial Supply Chain and Constructing of Digital CBD	149
Research on the Supply Chain Partnership Relations with Association Rule Analysis	154

Studies on Horizontal Competition among Homogenous Retailers Based on Agent-based Approach	158
Study of Information Supply Chain and Artificial Neural Network's Related Application	164
Study on Distribution Patterns of E-Business Based Agricultural Products Logistics	169
Study on Supply Chain Management in Electronic Commerce.....	175
Supply and Demand Management under Inducement of Price Discounts – A Monte Carlo Simulation Analysis	180
Taking Advantage of E-Logistics to Strengthen the Competitive Advantage of Enterprises in China.....	185
The Distribution Network Planning Using Geographic Information System (GIS).....	190
The Impact of IT Capability and E-Business on the Supply Chain Performance	193
The Implementation of an Intelligent Logistics Tracking System Utilizing RFID	199
The Operating Cost in a Tele-working Organization.....	205
The Virtual Service Value-Chain: Disruptive Technology Delivering Competitive Advantage for the Services Industry	211
The Web Services on Mobile Shopping Automation	219
Trust, Relationship Commitment and Supply Chain Integrations in China	223
A Comparison of Brand Loyalty in Online and Physical Channels: An Empirical Application to Supermarkets	229
A Defection Detection Procedure Using SOM and Markov Chain : A Case of On-Line Game Provider	233
A Global Comparative Study of Online Auction Markets	239
A Model of Customer Lifetime Value Consider with Word-of-mouth Marketing Value	245
An Exploratory Study of the Effects of CRM Practices on CRM Effectiveness and Business Performance	249
An Intelligent Model for Stock Investment with Buffett Strategy, Classifier System, Neural Network and Linear Programming	255
Analysis of Costs, Benefits and ROI of CRM Implementation	261
Comparing Sales Strategies Using the Markov Chain Relationship Model	266
Consumers' Trade-Offs of Online Auctions Alternatives -- An Application of Conjoint Analysis	272
Cross-Cultural Differences of Online Store Image Perception and Customer Purchasing Intention	278
E-Business Adoption in Marketing and its Relationship with Factors in SWOT Analysis: an Empirical Investigation of Small Software Businesses	284
Extreme Availability: Determining the Limits of Availability in Commercial Systems.....	290
Impact of Internet Banking on Customer Satisfaction and Loyalty: A Conceptual Model	298
Internet Shopping Behavior among Singaporeans: Gender and Educational Issues	304
Markov Chain-based Clustering Analysis of Customers and WebPages	309
Modeling Competitive Diffusion Process in Japanese ADSL Service Market.....	314
Neural Networks Based Integrated Evaluation Method for the Effectiveness of CRM	320
Successful Customer Relationship Management: Antecedents and Imperatives	325

Supporting Small Businesses' Penetration of New International Markets - Practitioner's Approach.	329
The Design of a Lean CRM Software Process Model	335
The Development of RFID in Healthcare in Taiwan	340
The Glance of an Observer on the War of Brands of Food Products	346
Theoretical Framework of Empirical Study on Consumer Behavior in Web-Based Commerce	355
Willingness to Pay for Digital Services: Challenges for Future Research	359
Working Mechanism and Structure of Customer Services Support System	364
A Component-based Framework for Distributed Business Simulations in E-Business Environments	370
A Global Perspective of Business Usage of Web Technology	376
A Personalized Recommender System Based on Explanation Facilities Using Collaborative Filtering	382
A Process Model for Assessing the Distribution Options for Horizontal E-Business Portals ...	388
An Agile Protocol for E-Commerce	393
B2B Electronic Markets: A Conceptual Model for Analyzing the Sources of Business Value..	397
Benefit Evaluation Model of E-Government Procurement	402
E-Business Value Process Modelling	408
Evolution of Service Model and Architecture of ASP Business	414
Integrated Optimization Model of Production Planning and Scheduling for Batch Production	419
Mapping from BPMN-Formed Business Processes to XPDL Business Processes	422
Methodological Approach for Choosing Portfolio of E-Business Models	428
Rationality Validation of Business Process Model by Simulation Method	435
Research on Mobile Agent-based E-Commerce System Framework	440
Towards Collaborative Travel Recommender Systems	445
Web Service Based Universal Management of Workflow Resources	452
A Comparative Analysis of International and Chinese Electronic Commerce Research	460
A Conceptual Framework for E-tailing Quality and Its Evaluation	466
A Multi-Agent System for E-Business Processes Monitoring in a Web-Based Environment ...	470
A Study of Competences and Indicators for Electronic Commerce Professional Managers	476
A Study of Influencing Traditional Travel Agencies' Decision Making to Introducing B2C E-commerce	482
A Value Chain Based Analysis of Electronic Commerce's Transformation of China's Pharmaceutical Industry	488
An Exploratory Study of the Electronic Distribution of Hotel Rooms in Europe	493
An Initial Study on the Intensifying Enterprises Information Management	499
Antecedents of E-Business Assimilation in Manufacturing SMEs	505
Business Drivers for Application Servicing and a Software-as-a-Service Model	511
Comparison of E-Commerce Practices between Large Enterprises and Small to Medium Enterprises	517

Demonstrative Analysis of Hi-tech Enterprises Technology Innovation Capability Based on ISM	522
Determinants of E-Cmmerce Adoption in SMEs	527
E-business Management System Based on Coordinated Center for Dealer	533
E-Business Oriented Optimal Online Auction Design	538
Economies of Information Consumer Commodities – An Introduction to Conceptualising Forms of Information Capitalism by Two Cases	544
Effective and Strategic Use of Electronic Reverse Auctions	550
How E-Ready is E-Sri Lanka?	556
Internet Role in Lithuania SMEs internationalization	562
Investigating the Use of the Business Model Concept through Interviews	568
Logistical Performance as a Key Factor of Success for European Food E-tailers– Three Paradoxes and a Research Agenda	574
Research on Equipment and Spare Parts Management Based on Theory of Constraints	581
Semantic Innovation Management	586
The Analysis of the Difficult Points on Developing E-Commerce of the Western Region in China	591
The Business Process Investigation in the Perspective of Customer Value	596
The Internet and the Organizational Adoption of Electronic Business: A Research Agenda	604
Towards a Definition of IT Management Sophistication in Small Firms	610
Towards a Framework of Enterprise Integration: Linking Quality Management, Organizational Learning and Information Technology	616
Towards Integrated Document Management in Networked Product Development	622
Transaction Governance Structures of Brick-and-Mortar and Click-and-Order: An Instrument Design Study	628
A Study on Design Process Integration between Marketing and R&D of Notebook	634
A Web-based Operation Management System for Distributed Divisional Organizations	640
Apply MDA to E-Business: MDA Based Workflow Solution as an Example	646
E-Alliances: Search and Match for Small Business Partnership	650
Exploring Sources of Competitive Advantages in E-business Application in Mainland Chinese Real Estate Industry	655
Research on Collaborative Commerce Model Based on Web Services	661
Research on Modular Business Processes	665
Research on Multi-Agent Based Information Management Platform for Enterprise Business Cooperation	669
Study of cooperative product business support platform for new product development	674
Supporting Collaborative Communication in a Multi-layer Meta-process Model for Evolutionary Shared Workflows	677
Supporting Collaborative Communication in a Multi-layer Meta-process Model for Evolutionary Shared Workflows	683
Visionary Management: Footstone of Sustained Firm Growth	689
A Risk Identification Method of Virtual Enterprise	693

A VOC Based Approach to Support Virtual Organizations	698
Refinement for Ontology Evolution in Virtual Enterprises	702
Risk Evaluation for Virtual Enterprise	708
Risk Sorting for Enterprise under EC Environments.....	714
The Research of Product Graphical Information Sharing Technology of Virtual Manufacturing Enterprise in E-Commerce Environment.....	719
Virtual Enterprises Risk Management DSS under Electronic Commerce.....	724
Virtual Manufacturing and Virtual Spaces	728
Web Services –An Enabling Technology for Trading Partners Community Virtual Integration	733
A Study of Behavioral Intention for Mobile Commerce Using Technology Acceptance Model	738
Ambient E-Services: Framework and Applications	743
An Extended Collaborative Filtering-based Recommendation Procedure for Multimedia Contents in M-Commerce	752
An Overview of the Impact of National Culture on the Adoption of Mobile Commerce	757
Diffusion of Mobile Phones in Taiwan: An Evaluation of Influence Sources	763
Explaining Intent of Adoption of Wireless Internet Services by Business Executives: The Case of Paris Airports’ Hotspots	769
Extracting Requirements for Ubiquitous Computing Technology-based IS Using Factor Analysis.....	774
Extracting Requirements for Ubiquitous Computing Technology-based IS Using Factor Analysis.....	777
IEEE 802.20 Based Broadband Railroad Digital Network – The Infrastructure for M-Commerce on the Train	780
Individual Telecommunications Tariffs in Chinese Communities: History as a Mirror of the Future, and Relevance for Mobile Service Development in China	786
Managing Critical Incidents in the Fine-Paper Supply Chain through Mobile Collaborative System.....	792
MOBICORS-Movie: A MOBILE COntents Recommender System for Movie.....	798
Mobile Payment System with Privacy Protection.....	804
Physicians’Opinions of an SMS Professional Medical News Service: Insights from an SMS Survey	810
Security Measures in Mobile Commerce: Problems and Solutions	816
Simulating Collaborative Mobile Services – An Approach to Evaluate a New Location Enabling Service.....	822
Supporting Salespersons CRM Efforts through Location Based Mobile Support Systems.....	829
Technology Attribute-Diffusion Pattern Nexus: Evidence from the Early Experiences of 3G Mobile Communications	836
The Implementation of Signing e-Document by Using the Wireless Identity Module in Cellular Phone	841
The Quest for Mobility: Designing Enterprise Application Framework for M-Business Practices	847
The Researches on the Mobile E-Business Implements and Mode in Small and Medium-sized Enterprise	853

Ubiquitous Computing Based Mobile Commerce : Toward the Ubiquitous Commerce	856
User-Centered Guidelines for Design of Mobile Applications	862
Using System Dynamics to Model Mobile Commerce Diffusion	867
From Control to Competition: Trust Institution in E-Business	873
Negative Reputation Rate as the Signal of Risk in Online Consumer-to-Consumer Transactions	877
Recent Trend of Ratings Business in Japan and Improving Proposal for Ratings Forecasting Model	883
Security Alert Management in E-Business Networks.....	889
The Impact of a ‘Theological’Web Interface on Trust in the Context of Electronic Commerce	895
Trust for E-Business Management	902
Trust Models in the E-Commerce Environment	910
A Managers’Guide to KMS Adoption and Diffusion	917
A Methodology to Support the Design and Deployment of Knowledge Management within Inter-organizational Networks.....	924
A Qualitative Approach to E-Learning ? A Dynamic Situation in Montpellier Business School.	930
Adaptive User Interfaces for Intelligent E-Learning: Issues and Trends.....	934
Computer-Mediated Communication and Innovation: Do Communication Media Properties Influence Innovative Thinking Processes?.....	944
Corporate Mission and Vision and Customer Knowledge Management for Increasing Competitive Advantages.....	950
Design and Development of an E-Learning Management System	956
Effective E-Teaching: The Case of Hong Kong	962
E-Learning Performance and Students’ Results Case of a French Business School	967
Electronic Disclosure and Financial Knowledge Management	978
From Idea to Knowledge –Generating New Knowledge on E-Business	984
Knowledge Management, as the Key Factor of Survival in New Competition Age	990
Study on the Cultivation of Core Competence Based on Tacit Knowledge.....	996
The Relationship between Knowledge Sharing Behavior and the Effectiveness of Information Systems Strategic Planning (ISSP): An Empirical Study in Taiwan	999
The Roots of the Uncertainty in the Enterprise Tech-innovation Process under the Net-Environment.....	1005
Transformation & Countermeasures on Business Operation Models in Uncertain Environment	1010
What All Stakeholders Should Do: Creating E-Learning Communities	1013
What E-Learning Providers and End Users Should Do Respectively before Initiating Such a Learning Project.....	1020
A High Performance XML Querying Architecture.....	1028
Application of Weighted Support Vector Machines to Network Intrusion Detection.....	1034
Customers Confidence in E-Business: An Evaluation of Australian Practices A Case Study	1039

Internet Security and Digital Certificates: How Much Do You Know About Them?	1044
Secure Mobile Agents in Electronic Commerce by Using Undetachable Signatures from Pairings	1047
Security Service Model for RFID Enabled Supply Chain.....	1053
Student Recognition and Awareness of Information Security in Course Learning from Management Information Systems and Computer Science Classes: An Empirical Investigation	1058
A Mediator to Integrate Databases and Legacy Systems: The Mediated Data Integration (MeDInt) Mediator	1063
A Multi-Attribute Group Decision Approach Based on Rough Set Theory and Application in Supply Chain Partner Selection.....	1069
A New Clustering Algorithm for Categorical Attributes	1074
A Personalized Commodities Recommendation Procedure and Algorithm Based on Association Rule Mining	1079
A Practical Approach to Credit Scoring.....	1084
A Research on Chaos Forecast with the Capital Market Structure Model	1090
A Three-phased Online Association Rule Mining Approach for Diverse Mining Requests ...	1094
A User Preference Classification Method in Information Recommendation System.....	1100
A Web-Based Decision Support Systems for the Chronic Deteriorating Diseases	1106
An Ant-based Clustering Algorithm in Data Mining	1110
An Interactive Web Meeting System Using Multicast Technology	1115
An Investment Decision Support System (IDSS) for Identifying Positive, Neutral and Negative Investment Opportunity Ranges with Risk Control in Stock Markets	1118
Bayesian Network Induction with Incomplete Private Data	1128
Constructing Control Process for Wafer Defects Using Data Mining Technique	1134
Decision Making and Analysis for Unexpected Road Blockages	1139
Discovering Fuzzy Functional Dependencies as Semantic Knowledge in Large Databases	1145
E-Business Data Warehouse Design and Implementation	1149
Forecasting the Consumption for Electricity in Taiwan	1152
Logistics Forecasting Using Improved Fuzzy Neural Networks System.....	1156
Maximising Technology Efficiencies for SMEs Using Computer Intelligence	1160
Mining the Change of Events in Environmental Scanning for Decision Support	1165
Monitoring the Supply of Products in a Supply Chain Environment: a Fuzzy Neural Approach	1171
PATSEEK: Content Based Image Retrieval System for Patent Database	1176
Privacy-Preserving Collaborative Association Rule Mining	1181
Privacy-Preserving Data Mining In Electronic Surveys	1188
Proposal of Improving Model for Default Probability Prediction with Logit Model on Non-Compensatory Rule.....	1195
Study of SOA Component Dynamic Scheduling Based on Mobile Agent Coalition.....	1202
Temporal Association Rule Mining in China's Closed-end Fund Data.....	1206
The Application of RFID in Emergency Medicine	1210

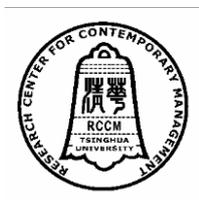
The Design of a Web Document Snapshots Delivery System	1216
Using Intelligent Agents to Build E-Business Software	1222
What Data Is Necessary To Data Mine For Knowledge?	1228
A User-Oriented Contents Recommendation System in Peer-to-Peer Architecture	1233
An Architecture of FoaF-based Peer 2 Peer Knowledge Management System	1239
Distributed Load Balancing in Peer-to-Peer Computing	1244
Survey of Security in Grid Services	1250
A Study of Internet Book Sale	1255
Achievements and Future Trends of E-Government Service – Implications from the SARS Outbreak	1260
Analysis of Price Dispersion in E-Market: A Case Study	1265
Current State of China’s Online Gaming Industry and the Obstacles in Development	1270
Developing a B2B E-Commerce Evolution Model -- The Case Studies of Seven Industries in Taiwan	1274
Digital Divide: Adoption of Internet in China	1279
E-business and Organizational Learning: A Case Study of Three Globally Operating Companies	1286
E-Business Evaluation in Australia: A Discussion Of Some Preliminary Findings.....	1291
E-Business: Factors of EAI Adoption in an Australian Electricity Company	1298
Electronic Business in Germany: Current Challenges and Future Perspectives - Results of an Explorative Survey -	1304
Exploring the Reality of E-Commerce Benefits among South African Businesses.....	1310
Intended Usage of Online Supermarkets: The Singapore Case	1317
Positioning of Firms on an E-Business Value Chain: the Case of Video on Demand.....	1322
Strategy and Risk Analysis of the ERP Projects in Three Chinese Cigarette Enterprises.....	1328
Towards a Knowledge Based Society: A Case Study on Entrepreneurship in Malaysia	1334
Analysis of Trademarks Act of Thailand and a Case Study: Conflict of Domain Name and Trademarks	1336
Are E-Bank Related Business Methods ^[1] Patent Subject Matter under Chinese Law?	1341
Continuous Audit Agent System	1348
Jurisdiction and Consumer Contract in E-Business	1353
Jurisdiction over E Commerce.....	1359
On the Novel Network Forensics Perspective of Enhanced E-Business Security	1364
Regulating Copyrights in Cyberspace	1370
Research on Policies to Support Telecommuting in China	1377
The New Vigilantism: Combating the Piracy of Copyrighted Materials on Peer-to-Peer Networks	1383
The Security Flaw of an Untraceable Signature Scheme	1388

The Fourth International Conference on Electronic Business

Dec 5 ~ 9, 2004

Beijing, China

PROGRAMME



SUPPORTED BY:

National Natural Science Foundation of China
Science Fund for Creative Research Group

HOSTED BY:

School of Economics and Management, Tsinghua University, Beijing, China
Research Center for Contemporary Management, Tsinghua University, Beijing, China

The Fourth International Conference on Electronic Business

Dec 5 ~ 9, 2004, Xiyuan Hotel, Beijing, China

Program Layout

SUNDAY, DECEMBER 5, 2004				
Ω	14:00-18:00	Registration (The Lobby of Xiyuan Hotel) * Registration will also be arranged on Dec 6, 8:00-12:00 am.		
Ω	19:00-20:00	Reception (Revolving Restaurant)		
MONDAY, DECEMBER 6, 2004				
Ω	9:00-12:00	Plenary Session (the Grand Banquet Room)		
Ω	12:00-13:00	Lunch (Food Street)		
Ω	13:30-15:35	Session M3-R1	Session M3-R2	Session M3-R3
Ω	15:35-15:55	Coffee Break		
Ω	15:55-18:00	Session M4-R1	Session M4-R2	Session M4-R3
TUESDAY, DECEMBER 7, 2004				
Ω	08:30-10:10	Session T1-R1	Session T1-R2	Session T1-R3
Ω	10:10-10:30	Coffee Break		
Ω	10:30-12:10	Session T2-R1	Session T2-R2	Session T2-R3
Ω	12:10-13:05	Lunch (Food Street)		
Ω	13:30-15:35	Session T3-R1	Session T3-R2	Session T3-R3
Ω	15:35-15:55	Coffee Break		
Ω	15:55-18:00	Session T4-R1	Session T4-R2	Session T4-R3
WEDNESDAY, DECEMBER 8, 2004				
Ω	08:30-10:10	Session W1-R1	Session W1-R2	Session W1-R3
Ω	10:10-10:30	Coffee Break		
Ω	10:30-12:10	Session W2-R1	Session W2-R2	Session W2-R3
Ω	12:10-13:05	Lunch (Food Street)		
Ω	13:30-15:35	Session W3-R1	Session W3-R2	Session W3-R3
Ω	15:35-15:55	Coffee Break		
Ω	15:55-18:00	Session W4-R1	Session W4-R2	Session W4-R3
Ω	19:00-20:00	Banquet (the Grand Banquet Room)		
THURSDAY, DECEMBER 9, 2004				
08:30: One day tour (OPTIONAL) Registration fee doesn't include the cost for one day tour, participants who plan to join the one day tour need to pay on site. Hotel organize one day trip every day. If you want to join the one day trip, please tell Xiyuan Hotel staff about your one day tour plan when you check in, Hotel staff will guide you to purchase the one day tour ticket. Details about payment and tourist information see Tour Information on our website. line 1: the Forbidden City, the Summer Palace and the Temple of Heaven line 2: The Great Wall, Ming Tombs				

The Fourth International Conference on Electronic Business

Dec 5 ~ 9, 2004, Xiyuan Hotel, Beijing, China

Program

SUNDAY, DECEMBER 5, 2004

Ω 14:00-18:00 Registration (The lobby of Xiyuan Hotel)

*Registration will also be arranged on Dec 6 8:00-12:00 am.

MONDAY, DECEMBER 6, 2004

Plenary Session

Chair: Jian Chen

Ω 9:00-10:00 Massive Quasi-Experimental Research Methods and the Discovery
of E-Business Knowledge

Robert J. Kauffman

Ω 10:00-10:30 Coffee Break

Chair: Jian Chen

Ω 10:30-11:30 Push-Pull Production: A Model for the Internet Age

David D. Yao

Ω 12:00-13:00 Lunch

Session M3-R1: SCM (I)

Ω 13:30-15:35 **On Inventory Strategies of Online Retailers**
Frank Y. Chen, S. H. Hum, Cheryl H. Sim

The Operating Cost in a Tele-working Organization
Xiaojun Li, Pengzhu Zhang, Fanghsin Liu

**A Study of Benefits of Information Sharing between Production Systems
and Service Centers**
N.Viswanadham, Vijay Desai

**Supply and Demand Management under Inducement of Price Discounts - A
Monte Carlo Simulation Analysis**
Henry C. Co

Coordination Mechanisms in the Supply Chain based on E-business
Suicheng Li, Xianxia Cui, Gongqian Liang

Session M3-R2: Case Study: Regional e-Business

Ω 13:30-15:35 **E-business and Organizational Learning: A Case Study of Three Globally
Operating Companies**
Marjatta Maula

Intended Usage of Online Supermarkets: The Singapore Case
Hui Tak Kee, David Wan

E-Business: Factors of EAI Adoption in an Australian Electricity Company
Ying Chen, Alex Wong

Exploring the Reality of E-Commerce Benefits among South African Businesses
Alemayehu Molla

Forecasting the Consumption for Electricity in Taiwan
Hsiaotien Pao, Tenpao Lee

Session M3-R3: E-readiness and Regional E-business Strategy (I)

- Ω 13:30-15:35 **The Analysis of the Difficult Point on Developing E-Commerce of the Western Region in China**
Luming Yang, Hongli Hu, Cuiping Zhu, Zhuqing Yang
- An Initial Study on the Intensifying Enterprises Information Management**
Peiji Shao, Jingjing Li, Jing Yang, Jie Wan
- An Exploratory Study of the Electronic Distribution of Hotel Rooms in Europe**
David C Gilbert, Liz Leekelley, Dominic W Beveridge
- How E-Ready is E-Sri Lanka?**
Velauthapillai Dhayalan, Reggie Davidrajuh
- A Study of Competences and Indicators for Electronic Commerce Professional Managers**
Chihhung Wu, Wenchang Fang, Gaoliang Wang

Session M3-R4: E-Logistics and Virtual Enterprises

- Ω 13:30-15:35 **A Study of Corporate Managerial Module and Information System Orienting Agile Virtual Enterprises**
Tong Shu, Shou Chen
- The Virtual Service Value Chain: Disruptive Technology Delivering Competitive Advantage for the Services Industry**
John Hamilton
- E-Logistics Comparative Positioning Model: A Multi-National Enterprise Airline Study**
John Hamilton
- Trust, Relationship Commitment and Supply Chain Integration in China**
Baofeng Huo, Xiande Zhao, Jeff Hoi Yan Yeung
- The Implementation of an Intelligent Logistics Tracking System Utilizing RFID**
Chichun Lo, Wenchuan Hsieh, Litsung Huang

Ω 15:35-15:55 **Coffee Break**

Session M4-R1: SCM (II)

- Ω 15:55-18:00 **Research on the Supply Chain Partnership Relations with Association Rule Analysis**
Huiying Zhang, Dezhi Zhu, Jin Du, Xin Chen
- Development Stages of Supply Chain Management in E-Business Environment**
JeongWook Choi
- Collaborative Planning, Forecasting and Replenishment: Demand Planning**

in Supply Chain Management
Hokey Min, Wenbin Yu

A Framework to Support Coalition Formation in Supply Chain Collaboration
Chattrakul Sombatteera, Aditya Ghose, Peter Hyland

Design of an E-SCM System for an Auto-parts Companies
Jun Su Lee, Jae Byung Rhu, Ki Seong Kim, Myung Hyun Oh, Jong Yeong Lee, Byung Ho Jeong

Session M4-R2: Recommendation System in E-business

- Ω 15:55-18:00 **A Personalized Recommender System Based on Explanation Facilities Using Collaborative Filtering**
Do Hyun Ahn, Hee Ae Lee
- A User-oriented Contents Recommendation System in Peer-to-Peer Architecture**
Hyea kyeong Kim, Jae Kyeong Kim
- An Extended Collaborative Filtering-based Recommendation Procedure for Multimedia Contents in M-commerce**
MeeYeon Kang, YoonHo Cho, JaeKyeong Kim
- MOBICORS-Movie: A MOBILE COntents Recommender System for Movie**
JaeKyeong Kim, YoonHo Cho, SeungTae Kim
- Towards Collaborative Travel Recommender Systems**
Cane Wingki Leung, Stephen Chifai Chan, Korris Fulai Chung
- A Defection Detection Procedure Using SOM and Markov Chain: A case of On-line Game Providers**
Kyunghee Chae, Heeseok Song, Juchoel Choi

Session M4-R3: E-readiness and Regional E-business Strategy (II)

- Ω 15:55-18:00 **Digital Divide: Adoption of Internet in China**
Kanliang Wang, Qing Liu
- Electronic Business in Germany: Current Challenges and Future Perspectives-Results of an Explorative Survey**
Tomaso Forzi, Ediz Kiratli, Stefan Bleck
- E-Alliances: Search and Match for Small Business Partnership**
Alan Chonghua Guo
- A Comparative Analysis of International and Chinese Electronic Commerce Research**
Jun Yong Xiang, Jae Kyeong Kim, Hyouk O

Session M4-R4: E-business Model (I)

- Ω 15:55-18:00 **A Component-based Framework for Distributed Business Simulations in E-Business Environment**
Alexander Hampel, Edward W.N. Bernroider
- A Process Model For Assessing the Distribution Options For Horizontal E-Business Portals**
Tapati Bandopadhyay, Pradeep Kumar, Tapas Mahapatra
- E-business Value Process Modelling**
Baolin Wu, Li Li, Yun Yang

TUESDAY, DECEMBER 7, 2004

Session T1-R1: Customer Relation Management (I)

- Ω 08:30-10:10 **Collaboration Strategy of Business-to-Business Enhancing Customer Satisfaction**
Chulsoo Kim, Youn Tae Kim
- Customers Confidence in E-business: An Evaluation of Australian Practices A Case Study**
Daniel Chandran
- A Comparison of Brand Loyalty in Online and Physical Channels: An Empirical Application to Supermarkets**
Javier Cebollada
- An Exploratory Study of the Effects of CRM Practices on CRM Effectiveness and Business Performance**
Jashen Chen, Russell K.H. Ching, Eldon Y. Li, Yiling Liao

Session T1-R2: E-learning and Knowledge Management (I)

- Ω 08:30-10:10 **What E-learning Providers and End Users Should Do Respectively before Initiating Such a Learning Project**
Zheng Zhao
- Effective E-Teaching: The Case of Hong Kong**
Candy K. Y. Ho, Ivy Chan
- Adaptive User Interfaces for Intelligent E-Learning: Issues and Trends**
Abdul Rahim Ahmad, Otman Basir, Khaled Hassanein
- From Idea to Knowledge-Generating New Knowledge on E-Business**
Johanna Tommila, Anne Mari Järvelin, Marko Seppä

Session T1-R3: SME E-business Strategy (I)

- Ω 08:30-10:10 **Comparison of E-Commerce Practices between Large Enterprises and Small to Medium Enterprises**
Sehyung Cho, Yongkyun Chung, Seungchul Kim
- Determinants of E-Commerce Adoption in SMEs**
Pham Thi Thanh Hong, Nguyen B. Giang
- Towards a Definition of IT Management Sophistication in Small Firms**
Theekshana Suraweera, Paul Cragg, Annette Mills
- Maximising Technology Efficiencies for SMEs Using Computer Intelligence**
Adrian Collins

Session T1-R4: E-business Model (II)

- Ω 08:30-10:10 **Markov Chain-Based Clustering Analysis of Customers and WebPages**
ChangShou Deng, Pie Zheng, Yanling Yang, Bingyan Zhao
- Comparing Sales Strategies Using the Markov Chain Relationship Model**

Bocheng Chen, W.H. Ip, Peng Sheng, H.L. Yu
Business Drivers for Application Servicing and a Software-as-a-Service Model
Jiri Vorisek
An E-Logistics Model for Effective Collaborative Commerce
Pradeep Kumar, Tapas Mahapatra

Ω 10:10-10:30 Coffee Break

Session T2-R1: Customer Relation Management (II)

Ω 10:30-12:10 Analysis of Costs, Benefits and ROI of CRM Implementation
Lun Hou, Xiaowo Tang
A Web-Based Decision Support Systems for The Chronic Deteriorating Diseases
Chichang Chang
Internet Shopping Behavior among Singaporeans: Gender and Educational Issues
Hui Tak Kee, David Wan
Successful Customer Relationship Management: Antecedents and Imperatives
Frederick L. Capossela

Session T2-R2: E-learning and Knowledge Management (II)

Ω 10:30-12:10 A Methodology to Support the Design and Deployment of Knowledge Management within Inter-organizational Networks
Tomaso Forzi, Meikel Peters, Stefan Bleck
A Managers' Guide to KMS Adoption and Diffusion
Jun Xu, Mohammed Quaddus
E-Learning Performance and Students' Results Case of a French Business School
Emmanuel Houze, Regis Meissonier
A Qualitative Approach to E-Learning? A Dynamic Situation in Montpellier Business School
Dominique Drillon, Eleni Boursinou

Session T2-R3: SME E-business Strategy (II)

Ω 10:30-12:10 Supporting Small Businesses' Penetration of New International Markets - Practitioner's Approach
Jaroslav Jandos
E-Business Adoption in Marketing and its Relationship with Factors in SWOT Analysis: an Empirical Investigation of Small Software Businesses
Edward W.N. Bernroider, Alexander Hampel
Antecedents of E-Business Assimilation in Manufacturing SMEs
Louis Raymond, Francois Bergeron, Sam Blili
Internet role in Lithuania SMEs Internationalization
Rimantas Gatautis, Panagiotis Damaskopoulos

Session T2-R4: P2P Computing and XML

- Ω 10:30-12:10 **Distributed Load Balancing in Peer-to-Peer Computing**
 Shan Zhang, Zheng Qin
- An Architecture of FoaF-based Peer 2 Peer Knowledge Management System**
 Xuan Zhou, Jian Zhou, Qing Li, Yuliu Chen, Lars Ludwig
- A High Performance XML Querying Architecture**
 Fangju Wang, Hui Shen
- Extreme Availability:Determining the Limits of Availability in Commercial Systems**
 David W. Coleman, Todd W. Mummert

Ω 12:10-13:05 **Lunch**

Session T3-R1: Online Auction: Strategy and Practice

- Ω 13:30-15:35 **A Study of Single-vendor and Multiple-retailers Pricing-Ordering Strategy under Group-Buying Online Auction**
 J. Chen, Yunhui Liu, Xiping Song
- Bidding Strategies Analysis for Procurement Combinatorial Auctions**
 J. Chen, He Huang
- A Global Comparative Study of Online Auction Markets**
 James K. Ho
- Consumers'Trade-Offs of Online Auctions Alternatives - An Application of Conjoint Analysis**
 Hector Mingta Tsai, Yichen Li
- Effective and Strategic Use of Electronic Reverse Auctions**
 Chia Yao Lee, Brian Corbitt

Session T3-R2: E-learning and Knowledge Management (III)

- Ω 13:30-15:35 **Electronic Disclosure and Financial Knowledge Management**
 Jerome Yen, Percy Yuen, Belinna Bai
- Knowledge Management, As the Key Factor of Survival in New Competition Age**
 Masoud Karami Zarandi, Ehsan Ghaffari Cherati
- The Relationship between Knowledge Sharing Behavior and the Effectiveness of Information Systems Strategic Planning (ISSP): An Empirical Study in Taiwan**
 Jungchi Pai
- Computer-Mediated Communication and Innovation: Do Communication Media Properties Influence Innovative Thinking Processes?**
 Leif Jarle Gressgård, Gunnar E. Christensen
- A User Preference Classification Method in Information Recommendation System**
 Chentung Chen, Weishen Tai
- Corporate Mission and Vision and Customer Knowledge Management for Increasing Competitive Advantages**
 Hwanyann Su, Yichen Lin, Chiuhua Lai

Session T3-R3: Mobile Commerce (I)

- Ω 13:30-15:35 **Security Measures in Mobile Commerce: Problems and Solutions**

Sanwar Ali, Waleed Farag, Mohammad A. Rob

Ubiquitous Computing Based Mobile Commerce: Toward the Ubiquitous Commerce

Sangjun Kim, Euiho Suh, Keedong Yoo

Simulating Collaborative Mobile Services - An Approach to Evaluate a New Location Enabling Service

Peter Dornbusch, Florian Fuchs, Martin Huber

Mobile Payment System with Privacy Protection

Kuenliang Sue, Wanpu Hsieh

The Implementation of Signing E-Document by Using the Wireless Identity Module in Cellular Phone

Chengyuan Ku, Yenfang Ho, Yiwen Chang

Session T3-R4: E-business Model (III)

Ω 13:30-15:35 A Model of Customer Lifetime Value Consider with Word-of-mouth Marketing Value

Yongmei Liu, Zhenjun Zhao, Yanlong Zhang, Hongjian Lu

Logistical Performance as a Key Factor of Success for European Food E-tailers—Three Paradoxes and a Research Agenda

Gilles Pache, Dominique Bonet

A Study of Influencing Traditional Travel Agencies' Decision Making to Introducing B2C E-commerce

Chenyi Hsu, Yunghsin Wang

Exploring Sources of Competitive Advantages in E-business Application in Mainland Chinese Real Estate Industry

Yan Tao, C. Matt Hinton

E-Business Evaluation In Australia: A Discussion Of Some Preliminary Findings

Mohini Singh, John Byrne

Ω 15:35-15:55 Coffee Break

Session T4-R1: Business Process Innovation

Ω 15:55-18:00 A Study on Design Process Integration Between Marketing and R&D of Notebook

Linlang Tang, Yubin Chiu, Hungsheng Lin

On-demand Business Process Integration Based on Intelligent Web services

Xiaohua Lu, Yinsheng Li, Ying Huang

E-Booking Control Problem in Sea Cargo Logistics

Chengxuan Cao, James Soo Keng Ang, HengQing Ye

The Impact of IT Capability and E-business on the Supply Chain Performance

LingLang Tang, ChiaChi Shih, ChengChuang Hon

Mapping from BPMN-Formed Business Processes to XPD L Business Processes

Moonyoung Jung, Hak Soo Kim, Myung Hyun Jo, Kyung Hyun Tak, Hyun Suk Cha, Jin Hyun Son

Session T4-R2: Case Study: Online Business Activities

- Ω 15:50-17:55 **Analysis of Price Dispersion in E-Market: A Case Study**
Zhiyuan Ge, Yong Shao
- Strategy and Risk Analysis of the ERP Projects in Three Chinese Cigarette Enterprises**
Yan Zhu, J. Chen
- A Study of Internet Book Sale**
Gaojia Jin, Takasi Nemoto
- Towards a Knowledge Based Society: A Case Study on Entrepreneurship in Malaysia**
Goh Hong Wee, Goh Peck Chen
- Economies of Information Consumer Commodities - An Introduction to Conceptualising Forms of Information Capitalism by Two Cases**
Jussi Okkonen, Pasi Virtanen

Session T4-R3: Mobile Commerce (II)

- Ω 15:55-18:00 **User-Centered Guidelines for Design of Mobile Applications**
Xiaowen Fang, Susy Chan, Jacek Brzezinski, Shuang Xu, Jean Lam
- Individual Telecommunications Tariffs in Chinese Communities: History as a Mirror of the Future, and Relevance for Mobile Service Development in China**
H. Chen, L-F Pau
- Supporting Salespersons CRM Efforts through Location Based Mobile Support Systems**
Chihab BenMoussa
- Managing Critical Incidents in the Fine-Paper Supply Chain through Mobile Collaborative System**
Vaida Kadyte
- An Overview of the Impact of National Culture on the Adoption of Mobile Commerce**
Nasrin Rahmati

Session T4-R4: E-business Model (IV)

- Ω 15:55-18:00 **Study on Distribution Patterns of E-Business Based Agricultural Products Logistics**
Yingrong Lu, Yinsheng Yang, Feng Lu, Xu Sun
- Towards Integrated Document Management in Networked Product Development**
Hannu Laesvuori, Paavo Kotinurmi
- Current State of China's Online Gaming Industry and the Obstacles in Development**
Bin Qiu, Shaoqin Sun, Cunyan Fan, Shujen Chen
- Positioning of Firms on an E-business Value Chain : The Case of Video on Demand**
Nabyla Daidj, Olivier Epinette, Pierre Vialle

WEDNESDAY, DECEMBER 8, 2004

Session W1-R1: Data Mining and Data Warehouse

- Ω 08:30-10:10 **What Data Is Necessary To Data Mine For Knowledge?**
Paul Alpar
- E-Business Data Warehouse Design and Implementation**
Xudong Chen
- Constructing Control Process for Wafer Defects Using Data Mining Technique**
Leeing Tong, Hsingyin Lee, Chifeng Huang, Changke Lin, Chienhui Yang
- Mining the Change of Events in Environmental Scanning for Decision Support**
Mengjung Shih, Duenren Liu, Churnjung Liao, Chinhai Lai

Session W1-R2: E-business Policy and Cyber Law

- Ω 08:30-10:10 **Are E-bank Related Business Methods Patent Subject Matter under Chinese Law?**
Yongjun Jin
- Research on Policies to Support Telecommuting in China**
Peipei Wang, Pengzhu Zhang, Fangfang Xu
- Regulating Copyrights in Cyberspace**
Sylvia Mercado Kierkegaard
- Jurisdiction and Consumer Contract in E-Business**
Zheng Tang

Session W1-R3: Workflow Management

- Ω 08:30-10:10 **Manufacturing Enterprises Value Chain Model Analysis within the Context of the Network Economy**
Xuejun Xu, Li Luo, Xuelan Hua, Jianhui Huang
- A Two Phase Verification Algorithm for Cyclic Workflow Graphs**
Yongsun Choi
- Apply MDA to E-Business: MDA Based Workflow Solution as an Example**
Jiawei Lee, Jungsing Jwo
- Supporting Collaborative Communication in a Multi-layer Meta-process Model for Evolutionary Shared Workflows**
Damien Bright, Gerald Quirchmayr, Roger Tagg

Session W1-R4: Security in E-business

- Ω 08:30-10:10 **Security Alert Management in E-Business Networks**
Allan Lam, Pradeep K. Ray
- Internet Security and Digital Certificates: How Much Do You Know About Them?**
Nena Lim
- Trust for E-Business Management**
Liaquat Hossain, Rolf T. Wigand
- Student Recognition and Awareness of Information Security in Course Learning from Management Information Systems and Computer Science Classes: An Empirical Investigation**
Qinyu Liao, Xin Luo, Kirk Arnett

Ω 10:10-10:30 **Coffee Break**

Session W2-R1: E-business Management Systems

- Ω 10:30-12:10 **E-business Management System Based on Coordinated Center for Dealer**
Ding Zhang, Jingwen An, Zhongqing Cheng
- The Business Process Investigation in the Perspective of Customer Value**
Ray Tsaih, Wanying Lin, Miaoju Hung, Yaling Cheng
- Towards a Framework of Enterprise Integration: Linking Quality Management, Organizational Learning and Information Technology**
Jiunn Chieh Lee
- Investigating the Use of the Business Model Concept through Interviews**
Alexander Osterwalder, Yves Pigneur

Session W2-R2: Virtual Organization

- Ω 10:30-12:10 **Virtual Manufacturing and Virtual Spaces**
Mélissa Saadoun
- A Risk Identification Method of Virtual Enterprise**
Min Huang, Fei Xu, Xing wei Wang, W.H. Ip, K.L. Yung
- A VOC Based Approach to Support Virtual Organizations**
Hua Zhou, Min Liu, Cheng Wu
- Refinement for Ontology Evolution in Virtual Enterprises**
Li Li, Baolin Wu, Yun Yang

Session W2-R3: Decision Support Methods

- Ω 10:30-12:10 **A Multi-attribute Group Decision Approach Based on Rough Set Theory and Application in Supply Chain Partner Selection**
Wenqi Jiang, Zhongsheng Hua
- Discovering Fuzzy Functional Dependencies as Semantic Knowledge in Large Databases**
Xue Wang, Guoqing Chen
- An Intelligent Model for Stock Investment with Buffett Strategy, Classifier System, Neural Network and Linear Programming**
Anping Chen, Weili Lin, Yenchu Chen
- An Investment Decision Support System (IDSS) for Identifying Positive, Neutral and Negative Investment Opportunity Ranges with Risk Control in Stock Markets**
Wanli Chen

Session W2-R4: Reputation Management

- Ω 10:30-12:10 **Willingness to Pay for Digital Services: Challenges for Future Research**
Herbjorn Nysveen, Per E. Pedersen
- Negative Reputation Rate as the Signal of Risk in Online Consumer-to-consumer Transactions**
Dahui Li, Zhangxi Lin
- The Impact of a 'Theological' Web Interface on Trust in the Context of Electronic Commerce**
Haytham Siala, Feroz Siddiqui

A Practical Approach to Credit Scoring

Jae H. Min, Youngchan Lee

**Feedback as an Indicator of Reputation: Interpreting and Leaving
Feedback by Buyers in Online Auctions**

Sophia Wang, Hsiuching Chiu, Alice Hiuying Hon

Ω 12:10-13:05 **Lunch**

Session W3-R1: Privacy and Patent Protection in E-business

Ω 13:30-15:35 **PATSEEK: Content Based Image Retrieval System For Patent Database**
Avinash Tiwari, Veena Bansal

**The New Vigilantism: Combating the Piracy of Copyrighted Materials On
Peer-To-Peer Networks**

Alvin Stauber

The Security Flaw of an Untraceable Signature Scheme

Chinchen Chang, Yafen Chang

**Analysis of Trademarks Act of Thailand and a Case Study: Conflict of
Domain Name and Trademarks**

Settapong Malisuwan

Jurisdiction over E Commerce

Bill Loving, Xiangping Liu

Session W3-R2: Web-service Modeling and Applications

Ω 13:30-15:35 **An XCS-Based Intelligent Searching Model for Cross-Organization Identity
Management in Web Service**
ChiChun Lo, Weili Lin, Jay Wu

The Web Services on Mobile Shopping Automation

Chengchih Yang, Chengmin Lin, Shihung Chen

Research on Collaborative Commerce Model Based on Web Services

Liwen Wu, Yingqiang Duan

Business Interactions with Web Services

Youcef Baghdadi

Shopbot, Market Structure, and Internet Price Dispersion

Xia Chen, Jianan Wu

Session W3-R3: Mobile Commerce (III)

Ω 13:30-15:35 **The Quest for Mobility: Designing Enterprise Application Framework for
M-Business Practices**
Chianhsueng Chao

Using System Dynamics to Model Mobile Commerce Diffusion

Wenqing Wang, France Cheong

**Technology Attribute-Diffusion Pattern Nexus: Evidence from the Early
Experiences of 3G Mobile Communications**

Nir Kshetri

**Extracting Requirements for Ubiquitous Computing Technology-based IS
Using Factor Analysis**

Natalia Em, Euiho Suh, Keedong Yoo

Diffusion of Mobile Phones in Taiwan: An Evaluation of Influence Sources
Hersen Doong, Yicheng Ku

Session W3-R4: Knowledge Discovery

- Ω 13:30-15:35 **A Personalized Commodities Recommendation Procedure and Algorithm Based on Association Rule Mining**
Jianyi Zhang, Yunfeng Wang, Jie Li
- Temporal Association Rule Mining in China's Closed-end Fund Data**
Ying Liu, Guoqing Chen
- Privacy-Preserving Collaborative Association Rule Mining**
Justin Zhan, Stan Matwin, Nathalie Japkowicz, Liwu Chang
- Semantic Innovation Management**
Jianqiang Li, David O'Sullivan
- A Three-phased Online Association Rule Mining Approach for Diverse Mining Requests**
Chingyao Wang, Shianshyong Tseng, Tzungpei Hong, Yianshu Chu

Ω 15:35-15:55 **Coffee Break**

Session W4-R1: Web-service Techniques

- Ω 15:55-18:00 **The Design of a Web Document Snapshots Delivery System**
David Chao
- Transaction Governance Structure of Brick-and-Mortar and Click-and-Order: An Instrument Design and Experimental Study**
Aimao Zhang
- A Mediator to Integrate Databases and Legacy Systems: The Mediated Data Integration (MeDIInt) Mediator**
Chaiyaporn Chirathamjaree
- An Interactive Web Meeting System Using Multicast Technology**
Ook Lee
- The development of RFID in Healthcare in Taiwan**
Frank Kuo, Yuehching Lee, Chuan Yi Tang
- The Application of RFID in Emergency Medicine**
Tsai WC, Frank Kuo, Yuehching Lee, F. Wu

Session W4-R2: Intelligent Agent Techniques and Applications

- Ω 15:55-18:00 **Continuous Audit Agent System**
Charles L. Y. Chou, Timon C. Du
- Using Intelligent Agents to Build E-Business Software**
Adrien Coyette, Manuel Kolp, Stéphane Faulkner
- Research on Multi-Agent Based Information Management Platform for Enterprise Business Cooperation**
Haigang Li
- A Web-based Operation Management System for Distributed Divisional Organizations**
Wei Shang, Yijun Li, Qiang Ye, Wenjun Sun

Session W4-R3: Mobile Commerce (IV)

- Ω 15:55-18:00 **IEEE 802.20 Based Broadband Railroad Digital Network – The Infrastructure for M-Commerce on the Train**
Fumin Zou, Xinhua Jiang, Zhangxi Lin
- Explaining Intent of Adoption of Wireless Internet Services by Business Executives: The Case of Paris Airports' Hotspots**
Madeleine Besson, Olivier Epinette, Olivier Segard, Pierre Vialle
- Ambient e-Services: Framework and Applications**
Yuanchu Hwang, Soetsyr Yuan
- A Study of Behavioral Intention for Mobile Commerce Using Technology Acceptance Model**
Yuchung Hung, Haoerl Yang, Chinghung Hsiao, Yunlu Yang
- Physicians' Opinions of an SMS Professional Medical News Service: Insights from an SMS Survey**
Shengnan Han, Pekka Mustonen, Matti Seppanen, Markku Kallio

Session W4-R4: E-service: Model and Applications

- Ω 15:55-18:00 **A Global Perspective of Business Usage of Web Technology**
Chang Liu, Wayne Huang
- B2B Electronic Markets: A Conceptual Model for Analyzing the Sources of Business Value**
Chia Yao Lee, Brian Corbitt
- Developing a B2B E-commerce Evolution Model - The Case Studies of Seven Industries in Taiwan**
Vincent Ming, Guohong Chang, Chuehju Wu, Chengying Lin
- Modeling Competitive Diffusion Process in Japanese ADSL Service Market**
Nozomi Nakajima, Jian Chen
- Evolution of Service Model and Architecture of ASP Business**
Kang Rae Cho, Ho Joon Choi
- Proposal of Improving Model for Default Probability Prediction With Logit Model on Non-Compensatory Rule**
Yoshikazu Sakamaki

Ω 19:00-20:00 **Banquet**

THURSDAY, DECEMBER 9, 2004

One day tour (OPTIONAL)

Registration: the lobby of Xiyuan Hotel

Reception: Revolving Restaurant (on the 26th floor), 旋转餐厅

Banquet: the grand banquet room (Close to lobby) (燕乐堂=大宴会厅)

Lunches: Food Street, (Outside of Xiyuan hotel, two minutes walk) 食街

Plenary Talk: the Grand Banquet room (燕乐堂)

Sessions: Hongyun Hall 1 and 2 (4th follor), Hongyun Hall 3 and 6 (3rd follor) 鸿运厅1, 2, 3, 6

A Framework to Support Coalition Formation in Supply Chain Collaboration

Chattrakul Sombatheera, Aditya Ghose, Peter Hyland

Decision Systems Laboratory, School of IT and Computer Science,
University of Wollongong, NSW 2500, Australia
{cs50, aditya, phyland} @uow.edu.au

ABSTRACT

This paper proposes a framework for agents in globally collaborative supply chain by applying the concept of coalition formation, a cooperation game in game theory. This framework provides mutual benefits to every party involved buyers, sellers and logistics providers. It provides a common gateway that allows individual parties to locate the right partners, negotiate with them, and form coalition in the best possible ways. The framework is applicable to real world e-business models, including B2C, B2B, supply chain and logistics, SME, etc. We firstly discuss common needs existing in today e-business. We then discuss about our framework, i.e., negotiation protocol and decision mechanism.

Keywords: coalition formation, game theory, negotiation protocol, decision mechanism

1. INTRODUCTION

The current business environment is characterized by large complex supply chains that are often global in reach and that are highly adaptive, being frequently re-configured to respond to dynamic business contexts. It is widely recognized collaboration across the supply chain is a key prerequisite for supply chain efficiency. The effective deployment of information technology and systems, specially the Internet and the worldwide web, has made new modes of complex, dynamic yet effective collaboration possible across supply chains. Collaboration in supply chains can take various forms. Suppliers and buyers might collaborate to increase buying/selling power, to reduce logistics costs or to aggregate capacity. The problem of determining an optimal set of collaborative arrangements/agreements for a given firm can be complex. This research seeks to develop automated (or semi-automated) negotiation protocols as a basis for building decision support functionality for dealing with this complexity.

The coalition formation problem ([4], [9]) considers techniques and criteria that might be used by a collection of (rational) agents to decide how they might group together to improve individual or social utility. Coalitions are ubiquitous in real-life settings. The theoretical underpinnings of approaches to coalition formation lie in the literature on multi-player games in game theory ([9]). Players negotiate among themselves about payoffs to decide which coalition to join. In reality, it is more complex than that. Self-interested agents, operating in dynamic environment such as supply chain, are under heterogeneous constraints. Furthermore, each agent has its own strategies that increase or decrease the value of each constraint thus affects the decision making of agents. A simple example of this is one is using express mail service, which costs more, due to a deadline while another is using ordinary mail service, which costs less, with further deadline. While time is changing in dynamic environment, the

value of a constraint varies thus affecting the utility of the agent. In contrast to traditional coalition formation study, where the coalition value is predefined and thoroughly known among agents, an agent has to calculate, according to its constraints and strategies, for the coalition that would give it maximum utility. We believe that two of the key components of successful coalition formation of self-interested agents are, one, quickly negotiating with other agents and, two, selecting the best possible coalition. Each agent, bounded by its own constraints, may negotiate with others to form a coalition, which is likely to yield maximum benefit. Such a coalition, however, may not be formed due to the constraints. So the agent has to look for the next best possible coalition by consulting with its internal utility mechanism. Negotiation and decision must be done in a timely fashion. There are works in coalition formation that discuss the formation of buyers, sellers and LPs separately. This work provides a basic framework that allows thorough collaboration among agents in supply chain. It includes two important components: a negotiation protocol and a decision mechanism. The negotiation protocol allows agents to exchange necessary information before deciding which coalition to join. A coalition can be as complex as a coalition of buyers, sellers and LPs. **We restrict attention to traditional supply chain activities for the sake of simplicity. Our focus is primarily on material flows, but note that many of the concepts developed in this work could apply equally well to these other forms of collaboration.**

2. COLLABORATION IN THE SUPPLY CHAIN

As discussed earlier, our focus in this paper is to develop coalition formation techniques that are appropriate to large, complex supply chains. In our domain, we can conceive of three categories of actors/agents: *buyers*, *sellers* and *logistics providers (LPs)*. In this view, a logistics provider is any organization that provides transportation, warehousing

or other logistics-related services. A firm (buyer or seller) has a set of plans to produce goods. Each plan consists of multiple activities, which may be of two types: *external* and *internal*. External activities are those that involve with *i*) procuring raw materials from external suppliers and *ii*) transporting the raw materials from the suppliers' warehouses to appropriate manufacturing sites. Internal activities are those that deploy internal resources of the firms (e.g., in-house manufacturing, assembly etc.). These activities are often required to satisfy synchronization/scheduling constraints, and there are penalties associated with the violation of these constraints. In addition to these constraints, each agent/actor seeks to satisfy several (internal) objectives, e.g. buying the raw materials at the affordable prices, with acceptable quality etc. In modern web-enabled supply chains, the process of matching buyers with sellers (of goods or services) is often facilitated by online e-marketplaces. Our framework assumes the existence of multiple such e-markets, where sellers advertise their product or service offerings while buyers advertise their requirements. There are many different motivating factors that make firms collaborate. In other words, there are multiple distinct drivers for coalition formation.

The simplest form of a supply chain coalition is one where a buyer forms a coalition with a seller and an LP to satisfy one of the external activities in a plan (e.g. to supply and deliver a manufacturing input to a production process). Buyers sometimes form coalitions with other buyers for the purpose of buy-side aggregation, i.e., the aggregation of buying power. Sellers sometimes form coalitions with other sellers to aggregate selling power. Such coalitions are common for small sellers, such as in agricultural cooperatives (e.g. for micro-producers of dairy products). Logistics provider (LPs) may want to form coalitions to aggregate their service capacities. We note that we have only listed the basic drivers. Most real-life supply chain coalitions tend to have more than one motivating factor driving their formation. Thus a coalition consisting of several sellers and several LPs might be driven by the need to aggregate selling power (for the sellers), the need to aggregate production capabilities (again for the sellers) and the need to offer comprehensive production and delivery contracts (thus bringing in one or more LPs into the coalition). We also note agents/actors (i.e., buyers, sellers or LPs) could participate in more than one coalition at any given point in time. Thus a seller might participate in a coalition with other sellers to aggregate selling power, while simultaneously participating in a coalition with a different set of sellers in order to aggregate manufacturing capabilities.

Both quantitative and qualitative factors play a role in coalition formation decisions. Price, delivery cost and lead time are examples of the former while quality, satisfaction and trust are examples of the latter. In the current work, we focus on four key criteria - cost, lead

time, quality and trust - while acknowledging that other criteria might be relevant (though possibly less important). Cost is the sum of price and delivery cost. Lead time is the time elapsed between an order being placed and the good or service becoming available. Trust is usually taken to denote the degree of belief of an actor/agent that another actor/agent will live up to its commitments. These attributes are not orthogonal, and most decisions typically involve trade-offs between them. One may be willing to pay more for a shorter lead time (or higher quality or higher trust/reliability) and vice versa. One may, in some settings, be willing to accept longer lead times for the purpose of getting higher quality. Several other instances of such trade-off are common, but we do not list them all here. Although negotiation can be either bilateral or multilateral, we can in most instances reduce negotiation to the bilateral case without loss of generality. The simplest scenario is where single agents negotiate, e.g., a buyer with a seller and a buyer with an LP. A buyer can bilaterally negotiate with multiple sellers/LPs at the same time in order to find the best pair of a seller and an LP. The information exchanged is kept private to each agent. A more complex scenario is where multiple agents negotiate. An agent that stands to benefit the most from forming a coalition usually acts as the coalition leader. A coalition leader negotiates bilaterally with multiple agents of its own type (buyer, seller or LP) to establish a sectoral coalition, i.e., a coalition of buyers, or a coalition of sellers, or LPs etc. The object of such negotiation is usually to establish how the coalition value, i.e., the financial benefits of collaboration might be distributed across the members of the coalition.

3. A FRAMEWORK FOR SUPPLY CHAIN COALITION FORMATION

In our model, the agents are divided in three groups—the set $B = \{B_1, B_2, \dots, B_i\}$ of buyers, the set $S = \{S_1, S_2, \dots, S_j\}$ of sellers, and the set $L = \{L_1, L_2, \dots, L_k\}$ of LPs. The set $LOC = \{loc_1, loc_2, \dots, loc_m\}$ is a set of locations where sellers' manufacturing facilities or warehouses and buyers' sites are located. The set $G = \{G_1, G_2, \dots, G_n\}$ is a set of goods. Each good g is associated with a load belonging to the set $LD = \{LD_1, LD_2, \dots, LD_o\}$, where a load is defined by either weight or volume. The notion of a load as distinct from a good becomes relevant in the context of negotiation with LPs. The sets LOC , G and LD are common knowledge to all agents. We will assume that each agent (whether a buyer, seller or LP) has access to the following: 1) A history of the prior coalitions that the agent has participated in. 2) A history of what has transpired in the current round of negotiation. 3) An *operations management system*. In the instance of buyers and sellers, these would essentially be some form of production scheduling and optimization system. For LPs, these would be some form of logistics optimization system. The intent is to be able to query the system to obtain quotes that would be used in the process of

negotiation. Thus a seller should be able to query the operations management system by providing the product requirements and the date/time by which the product must be ready (typically a shipping deadline) to obtain a price quote. This quoted price is typically higher for early shipping deadlines (i.e. rush orders) and vice versa. Similarly, an LP should be able to query such a system with details on the loads, origin, destination and delivery deadlines to obtain a price quote to use in negotiation. Once again, rush orders are likely to result in higher price quotes and vice versa. An operations management system is assumed to have access to real-time data about the firm's current business context, and thus provides highly context-sensitive output. We do not discuss the design and implementation of such systems, but note that most firms (especially larger ones) tend to employ some version of these systems.

In the following, we present a simple negotiation protocol to support coalition formation. We first describe some basic assumptions made by the protocol. A message containing fields set to null is treated as a query. An agent replies to a query with an offer made by filling the null fields of the query with proposed values. Agents negotiate by modifying the values of the received offers and sending them back. Agents accept offers by sending acknowledge messages. When a deal is done, agents are obliged to discharge their commitments.

3.1 Buyer to Seller and Buyer to LP Negotiation

Direction	Message
B@S	<i>B2S(msg#, buyer, good, totalquantity, quality, (availabletime, origin, quantity), price, expirytime)</i>
B→S	<i>S2B(msg#, seller, good, quality, (availabletime, origin, quantity), price, expirytime)</i>
B@S	<i>B2Sack(msg#)</i>
B→S	<i>S2Back(msg#)</i>
B@L	<i>B2L(msg#, buyer, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)</i>
B→L	<i>L2B(msg#, LP, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)</i>
B@L	<i>B2Lack(msg#)</i>
B→L	<i>L2Back(msg#)</i>

Note: *B* is the buyer, *S* is the seller and *L* is the LP.

Table 1: Buyers' negotiation messages with sellers over buying goods and with LP over delivering goods.

The buyer begins negotiation by sending a message *B2S(msg#, buyer, good, quality, (availabletime, origin, quantity), price, expirytime)* to sellers asking for a quote for *totalquantity* units of the *good*. (*availabletime, origin, quantity*) is a list of triples, each of which specifies that a certain *quantity* of the good needs to be picked up from *origin* at *availabletime*. If the buyer is contacting the seller for the first time or wants an updated quote, *availabletime, origin, quantity* and *price* are set to null (i.e., the buyer does not wish to pre-specify an available-time, the exact location the good is to be sourced from, the quantity to be sourced from that location and the price the seller may wish to

offer). In subsequent messages from the buyer to the seller, the (*availabletime, origin, quantity*) triples may have non-null values. This is because the seller would have replied to the buyer with a similar list of tuples (see below) indicating that the seller would be able to meet the order by sourcing production from multiple locations with distinct available times. The *expirytime* provides the seller with a deadline within which it must reply – otherwise the request will be deemed to have expired. Upon receiving the message, each seller consults its operations management system, retrieves the price quote and available time, then replies to the buyer with the message *S2B(msg#, seller, good, quality, (availabletime, origin, quantity), price, expirytime)*, indicating that the *seller* will be able to offer the *good*, at a specific *price*, ready at *origin* on *availabletime*. The offer lasts for *expirytime* (in case the seller is not in a position to supply the *good*, it will not reply). If the buyer is satisfied with the offer, it will send the message *B2Sack(msg#)* back to the seller. Once the seller receives it, the deal is done—the buyer is obliged to pay money and the seller has to supply the good. If the buyer is not satisfied with the price, it negotiates on the price by sending the message *B2S(msg#, buyer, good, quality, (availabletime, origin, quantity), price, expirytime)* with the new price back to the seller again. If the seller is satisfied with it, it will send the message *S2Back(msg#)* back to buyer in order to seal the deal.

At the same time, the buyer also contacts one or more LPs to organize shipping of the product. It sends the message *B2L(msg#, buyer, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)* to LPs. The (*load, quantity, origin, destination, pickuptime, arrivaltime*) is a list of packaging tuples, each of which specifies that there are *quantity* of *loads* to be picked up from *origin* at *pickuptime* and to be delivered to *destination* at *arrivaltime*. For the first message from a buyer to an LP, or when the buyer seeks an updated quote, the *cost* will be set to null. Upon receiving the message, the LP consults its operations management system to compute the cost of the delivery job. It then sends the quote to the buyer with the message *L2B(msg#, LP, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)*. The quote will last for *expirytime*. If the buyer is satisfied with the quote, it finalises the deal by sending the message *B2Lack(msg#)* back to the LP. If the buyer seeks to make a counter-offer to an LP, it will set *load, quantity, origin, destination, pickuptime, arrivaltime* and *cost* to its desired values. It then sends a message *B2L(msg#, buyer, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)* to the LP. If the LP is happy with it, it will reply with the message *L2Back(msg#)*. Both buyer and LP are now contracted.

3.2 Decision Mechanism

At the top level, buyers have to obey a set of criteria in order to make decisions about their external activities.

We modify multi-attribute decision making scheme proposed by Keeney and Raiffa ([7]). Instead of having a fixed weight attached to each attribute, we propose to have a dynamic weight table for each attribute. The dynamic weight table associates a set of weights with a number of environment conditions in one-to-one manner. The environment conditions are the facts, which affect the weight, or importance, of an attribute. For example, the weight for time and price are equal to 1 in normal situation. In the face of a shortage of a raw material, the weight of time is 10 times more than that of the price. The environment conditions can be achieved from internal or external sources. They can be in form of relation among attributes and facts. Therefore, the near optimal is acceptable to experienced agents who know how to make good profit in the long run.

Strategic weight

Condition	Attribute	Weight
X_1	a_1	w_{11}
X_2	a_1	w_{12}
X_1	a_2	w_{21}
X_2	a_2	w_{22}

$$U = \sum_{q=1}^n w_{qp} a_q \text{ Where condition } X_p \text{ holds}$$

Buyer's ranking table

Sellers' offers	LPs' offers	$\sum_{q=1}^n a_q w_{qp}$
(msg ₁ , S ₁)	(msg ₇ , L ₇)	$\sum_{q=1}^n a_{q1} w_{q1} + \sum_{q=2}^n a_{q2} w_{q2}$
(msg ₂ , S ₁)	(msg ₈ , L ₈)	$\sum_{q=1}^n a_{q1} w_{q1} + \sum_{q=2}^n a_{q2} w_{q2}$
(msg ₃ , S ₂)	(msg ₉ , L ₈)	$\sum_{q=1}^n a_{q2} w_{q2} + \sum_{q=1}^n a_{q1} w_{q1}$

Figure 1: Strategic weight table, utility calculation and ranking for the best combination of seller and LP.

Each agent maintains the strategic weight table, which counsels the agent a value of the weight of an attribute under a certain condition. The weight of an attribute is in a range of integer numbers. The agent controls values of the lower and upper bounds. A condition change can be a stock shortage, a financial concern, a complaint from customer, or a combination of them. An agent learns from its experience that how the changing environment affects its business and how to react. The results of the reactions will be analysed by the agent and it will update the range of the weights and their values under such conditions. It specifies in the table about a certain condition X_p expressively. Each attribute a_q affected by the condition X_p will be assigned a specific weight w_{iq} to it. For example, when "there is a stock shortage of a raw material g"(good $g \in G$), the weight of the attribute "price" under this condition will be decreased while the weight of lead time will be increased. The agent updates its strategic weight table every time it senses a change in environment. This allows agent to act wisely in dynamic environment. Among the four attributes, trust is a special decision criterion in the sense that it depends on other attributes such as time and quality promised by other agents. Agents maintain trust differently. Each buyer has a trust table keeping record of other agents, including other buyers, sellers and LPs, and their trusted values. Each agent in the trust agent is associated with an integer value from range [-1,1]. The value is updated over times.

When a buyer is dealing with a new agent, it may give a low value to trust, e.g. 0.3 or 0.4 to the new agent. If the agent performs well, e.g., keeps up with schedule, provide high quality of good, etc., the buyer increases trust value for that agent. On the other hand, if an agent fails to keep up with its promise, the buyer decreases the trust value. All the offers made by sellers and LPs are kept in buyers' knowledge base. Whenever it needs to compute the utility, it checks for the current environment condition X_p and take the corresponding weight w_{pq} of each attribute a_q . It creates a new tuple ($msg\#, agent$) for each active offer and create a new ranking table composed of three fields: sellers' offer ($msg\#, agent$), LPs' offers ($msg\#, agent$) and $\sum_{q=1}^n a_q w_{qp}$. The first two fields are merely reference to the exact offers. The third field is the utility for each combination of sellers and LPs offers. The agent computes the utility for each combination by combining utility of the seller's offer and that of the LP's offer. The utility of an offer can be achieved by $\sum_{q=1}^n a_q w_{qp}$, where a_q is the attribute value retrieved from the corresponding offer, and w_{pq} is the weight of the attribute a_q under condition X_p .

3.3 Buyer to Buyer Negotiation

Direction	Message
$Bl @ Bm$	$Bl2Bm(msg\#, buyer, good, quality, (availabletime, origin, quantity), price, expirytime)$
$Bl \rightarrow Bm$	$Bm2Bl(msg\#, buyer, good, quality, (availabletime, origin, quantity), price, expirytime)$
$Bl @ Bm$	$Bl2BmDel(msg\#, buyer, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)$
$Bl \rightarrow Bm$	$Bm2BlDel(msg\#, buyer, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)$
$Bl @ Bm$	$Bl2BmAck(msg\#)$
$Bl \rightarrow Bm$	$Bm2BlAck(msg\#)$
$Bl @ Bm$	$Bl2BmDelAck(msg\#)$
$Bl \rightarrow Bm$	$Bm2BlDelAck(msg\#)$
Note: Bl is the leading buyer, Bm is a member buyer.	

Table 2: Buyers' negotiation messages on buying goods.

Buyers can form coalition to share discount on prices and delivery costs. A buyer can join any appropriate coalition. A buyer can join a coalition for discount on price and join another coalition for saving delivery cost. This allows agents to maximise benefits freely. When a buyer wants to form a coalition of buyers, it becomes a leading buyer, who tries to negotiate among buyers to form a buying coalition then negotiates on behalf of the coalition to agree on the prices with sellers. Firstly, the leading agent consults with its knowledge about other buyers who might be interested. Agents are considered being interested if *i*) they used to buy the same good at around this time previously, *ii*) they are located nearby, or *iii*) randomly selected. The leading agent sends the message $Bl2Bm(msg\#, buyer, good, quality, (availabletime, origin, quantity), price, expirytime)$ with value of *origin, quantity* and *price* set to null to all interested agents. The interested agents consider the invitation and if they find that they are in need of the same *good* of the same *quality* at *availabletime*, they will reply by sending the message $Bm2Bl(msg\#, buyer, good, quality, (availabletime, origin, quantity), price,$

expirytime) with the *quantity* set to an appropriate value back to the leading buyer. After the expiry time, the leading buyer assumes that all the truly interested buyers have responded. It then sums up all the *quantities* and sends the message *B2S(msg#, buyer, good, quality, (availabletime, origin, quantity), price, expirytime)* to sellers. The sellers will reply with the message *S2B(msg#, seller, good, quality, (availabletime, origin, quantity), price, expirytime)*.

Upon receiving quotes from sellers, the leading buyer select the one with maximal utility, computed the same way discussed above. It then computes the price for each interested agents. Then the leading agent sends the message *B12Bm(msg#, buyer, good, quality, (availabletime, origin, quantity), price, expirytime)* to interested agents. If the all interested agents agree, they send the message *Bm2BlAck(msg#)* and the coalition is formed. If any of them is not satisfied with the share, they may negotiate on the price by sending message *Bm2Bl(msg#, buyer, good, quality, (availabletime, origin, quantity), price, expirytime)*. The leading buyer may try to satisfy the demanding buyers by giving away its share on the discount (Kraus et al., 2004) up to a limit. Whenever it goes beyond that, the leading buyer send the message *B2S(msg#, buyer, good, quality, (availabletime, origin, quantity), price, expirytime)* with the price increased by the excess to the seller. If the seller accepts it, the leading buyer sends the *B12BmAck(msg#)* to all members. The buying coalition is now formed. All members are obliged to pay the bills. If the seller does not accept, the negotiation goes on, virtually similar to that between a single buyer to a single seller. If the time runs out, then all of them suffer loosing opportunity.

Any buyer can try to manage transportation for other agents. Those agents may have been already interested in buying within the same buying coalition or may be located nearby. The process is similar to forming buying coalition. The leading buyer sends message *B12BmDel(msg#, buyer, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)* with *load, quantity* and *cost* set to null to all interested agents. The recipient agents consider if the *origin, destination, pickuptime*, and *arrivaltime* are suitable with their plans. The interested agents sent message *Bm2BlDel(msg#, buyer, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)* with (*load, quantity, origin, destination, pickuptime, arrivaltime*) and *cost* set to appropriate values back to the leading buyer. The leading buyer collects all replies and send message *B2L(msg#, buyer, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)* with (*load, quantity, origin, destination, pickuptime, arrivaltime*) set to all members' requests and *cost* set to null to LPs. The LPs reply with message *L2B(msg#, LP, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)* with *cost* set to appropriate value. The leading buyer computes for

the best offer and compute for the cost for each interested agent. It then sends to interested agents the message *B12BmDel(msg#, buyer, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)* with *cost* set to their costs. If all the interested agents are happy with their costs, they reply with message *Bm2BlDelAck(msg#)*. The leading buyer sends the acknowledge message to the LP. The deal is done. If the interested agents are not satisfied with their costs, they negotiate with the leading agent by sending message *Bm2BlDel(msg#, buyer, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)* with their preferred *costs*. The leading agent can try to negotiate with the unhappy buyers by giving away its share on the saved cost up to a threshold. If the demanding goes beyond the threshold, the leading buyer negotiates with the LP. The process repeats backward and forward until every agent is satisfied or the time expires.

3.4 Seller to Seller and LP to LP Negotiation

Sellers can form coalition when appropriate. Once a seller receives a quote it considers out of its resources or considers outsourcing is beneficial, it negotiates with other sellers. It firstly consults with its own plan for its capability and consults with its knowledge base for outsourcing to other sellers. It computes for other sellers' share on goods and prices. It sends message *SI2Sm(msg#, seller, good, quality, availabletime, quantity, price, expirytime)* with price set to null to other sellers who can mutually cover the outsourcing. The recipients reply with the *prices*-quoted message *Sm2Sl(msg#, seller, good, quality, availabletime, quantity, price, expirytime)*. If the leading seller is satisfied with all quotes, its finalises the deals with other agents by sending message *SI2SmAck(msg#)*. If it is not satisfied with any of them, it can negotiate, similarly to how the leading buyer does it. Once all of the unsatisfied agents agree, they accept by sending message *Sm2SlAck(msg#)* back to the leader. Then the deal is done.

Direction	Message
<i>SI@Sm</i>	<i>SI2Sm(msg#, seller, good, quality, availabletime, quantity, price, expirytime)</i>
<i>SI-Sm</i>	<i>Sm2Sl(msg#, seller, good, quality, availabletime, quantity, price, expirytime)</i>
<i>Ll@Lm</i>	<i>Ll2Lm(msg#, LP, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)</i>
<i>Ll-Lm</i>	<i>Lm2Ll(msg#, LP, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)</i>
<i>SI@Sm</i>	<i>SI2SmAck(msg#)</i>
<i>SI-Sm</i>	<i>Sm2SlAck(msg#)</i>
<i>Ll@Lm</i>	<i>Ll2LmAck(msg#)</i>
<i>Ll-Lm</i>	<i>Lm2LlAck(msg#)</i>
Note: <i>SI</i> is the leading seller, <i>Sm</i> is a member seller. <i>Ll</i> is leading LP, <i>Lm</i> is member LP.	

Table 3: Sellers' negotiation messages over selling goods and LPs' negotiation messages over delivery jobs.

When contacted from either a buyer or a seller, an LP may form coalition with others when it does not have

enough resources or it finds that forming coalition is more profitable. The leading LP may divide whole distribution area into sections and decide to bilaterally negotiate with LPs, each of which is considered having enough potential to distribute the good on assigned section. The leading LP sends message $Ll2Lm(msg\#, LP, (load, quantity, origin, destination, pickuptime, arrivaltime), cost, expirytime)$ with $(load, quantity, origin, destination, pickuptime, arrivaltime)$ set to calculated values to potential agents. Each member agent tries to accommodate the on-negotiating job with its own resources and its on-going plans. If the member is satisfied with the offer, it sends message $Lm2LlAck(msg\#)$ back to the leading LP—the deal is done. If not, it negotiates with the leading LP similar to buyers and sellers. Some LPs may not be satisfied with bilateral negotiation, they may force the leading LP to manage the multilateral negotiation for the sake of fairness and efficiency. LP negotiation is concerned with the cost of doing the job and the profit. The cost is involved with the route a LP run regularly and the route for the new job.

4. RELATED WORK

Coalition formation is an active research area in multi-agent systems. Previous work has done some part of our proposing framework as the followings. Li et al. ([8]) addressed the combinatorial auction in coalition formation. Buyers submit their requests to the mediator who will allocate the good from sellers to them. Goldman et al. ([2]) searched for a strategy where sellers selected the most profitable deal while buyers looked for the most satisfiable sellers. Kraus et al. ([6]) proposed a compromise strategy for distributing profits among sellers in order to form coalition quickly. They found that agents who are willing to give away their profits actually earn more. Breban et al. ([1]) addressed the coalition formation based on trust among agents. Trust is used as a mechanism to enforce agents to commit themselves to the jobs as parts of the coalition. Hyodo et al. ([3]) addressed the optimal coalitions of buyers and sellers who are located in distributed sites. They deploy genetic algorithm to search for the optimal share over the discount. Klusch et al. ([5]) addressed problems in static coalition formation and proposed an algorithm for mobile agents to form coalition under dynamic environment. Agents can form overlapping coalitions.

5. CONCLUSION AND FUTURE WORK

We propose a simple framework, which involves negotiation protocol and decision mechanism. The negotiation protocol allows thorough communication, i.e., buyers to buyers, buyers to sellers, sellers to sellers,

buyer to LPs, and LPs to LPs. The decision mechanism is a modification of Keeney et al.'s ([7]), which allows agents to adapt its decision to suit with changing environment. Agents can select appropriate weights for the environment in order to make suitable decision. The traditional utility-based decision mechanism seems inadequate in complex settings in real world. New approaches, e.g., Markov decision processes, Bayesian networks, CPNets, are suitable for more complex and dynamic environment. Under uncertain environment, agents have to learn from the past and project it to forecast what will happen in the future. Efficient utilization of knowledge base should help agents perform better in dynamic coalition formation. Based on this framework, we want to develop agents who are able to evolve strategies in order to enhance cooperation. Instead of modeling supply chain competitive games, we want to explore that under changing environment, how agents can adapt their strategies to cooperate and be successful at the end.

REFERENCES

- [1] Breban, S. and J. Vassileva, "A coalition formation mechanism based on inter-agent trust relationships", *proceedings of the 1st AAMAS*, pp306-307, ACM Press, 2002.
- [2] Goldman, C. V., S. Kraus, and O. Shehory, "Agent strategies: for sellers to satisfy purchase-orders, for buyers to select sellers", *proceedings of the 10th European Workshop on Maamaw' 01*, Annecy, France, 2001.
- [3] Hyodo, M., T. Matsuo, and T. Ito, "An optimal coalition formation among buyer agents based on a genetic algorithm", *proceedings of IEA/AIE 2003*, pages 759-767, 2003.
- [4] Kahan, J. and A. Rapaport, *Theory of Coalition Formation*, Lawrence Erlbaum Associates, London, 1984
- [5] Klusch, M. and A. Gerber, "Dynamic coalition formation among rational agents". *Intelligent Systems*, IEEE, 17(3):42-47, 2002.
- [6] Kraus, S., O. Shehory, and G. Taase, "The advantages of compromising in coalition formation with incomplete information", *proceedings of AAMAS 04*, 2004.
- [7] Keeney, R. L. and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, John Wiley & Son, New York, 1976
- [8] Li, C. and K. Sycara, "Algorithm for combinatorial coalition formation and payoff division in an electronic marketplace", *proceedings of AAMAS 02*.
- [9] von Neumann, J. and O. Morgenstern, , 3rd edition, Princeton University Press, Princeton, 1953.

A Study of Benefits of Information Sharing between Production Systems and Service Centers

N. Viswanadham, Vijay Desai

The Logistics Institute –Asia Pacific, National University of Singapore
10 Kent Ridge Crescent, Singapore 119260
tlinv@nus.edu.sg; tlivd@nus.edu.sg

ABSTRACT

Integrated Manufacturing and Service Networks (IMSNs) are an assemblage of companies formed with the primary goal of delivering value in terms of both goods and services to the customer through out the product life cycle - from cradle to the grave. Manufacturing-service integration will lead to information sharing between design, manufacturing and sales units of the manufacturer and the service centres. While the integration between design and manufacturing is very well studied, the subject of collaboration between manufacturing units and service centre received practically no attention.

In this paper, we present two models each with increasing complexity to elucidate the benefits of integration and alert against problems like bullwhip effect percolating from supply chains into IMSN. The first one is a conceptual model that captures the relationship between manufacturing and service networks and makes a case for coordination between the two entities. The second model, IMSN with single product, evidences the benefits of collaboration in a more realistic setting.

In all the above models, we make use of cause effect diagrams and system dynamic models to study the benefits of integration. We study the effect of reactive (feedback) and also the proactive (feedforward) use of information for service centre resource management.

KeyWords: Manufacturing, Service, Bullwhip effect, Service Center Design.

1. INTRODUCTION

Integrated Manufacturing and Service Networks (IMSNs) are an assemblage of companies formed with the primary goal of delivering value in terms of both goods and services to the customer through out the product life cycle - from cradle to the grave. The product life cycle could be divided into design, manufacturing and service phases. During the design phase, collaborative teams design products taking into account the manufacturability, logistics requirements, ease in assembly, ease in fault diagnosis and the serviceability. The manufacturing phase involves procurement, manufacturing, distribution and sales. The service phase comprises of activities like preventive maintenance, testing and fault diagnosis, repair services, call centres and emergency services.

There is lot of interaction between the production phase and service phase, which has not been given attention. For example, a company which considers service phase in design and manufacturing phase can introduce onboard diagnostics for monitoring machine health and proactive maintenance to provide failure free performance. *It is estimated that reactive maintenance costs over six times more as compared to proactive maintenance and leads to increased customer satisfaction.* Expert system software combined with strategically located sensors and transducers (e.g., pressure, temperature, vibration, viscosity, wear debris,

and moisture) will provide comprehensive machine health monitoring and as a result making the whole process more efficient.

Every product sold is a potential source of revenue through sales of spare parts and after-sale services such as repair, maintenance and the upgrades. Consumers are demanding better service and OEMs are realizing the importance of providing good service. A recent GM study reported that \$9 billion in *after-sale revenue* yielded \$2 billion in profits. On the other hand, for producing \$2 billion profit from product sales GM will have to sell cars worth \$150 billion. In the sale of products and gaining customer loyalty, after sales service has become a significant factor. Decisions made during design and manufacturing phases affects the performance and cost of service phase. On the other hand, providing good service will boost the product sales. Feedback from service to design will help produce more reliable products. It is thus important to consider the maintenance, fault diagnosis, reparability and serviceability aspects in addition to manufacturability at the design stage.

With OEMs taking a more proactive role in service, they will have a better idea of customers needs. For example, most auto makers have relatively few opportunities to interact with car owners; most of the "touches" they do experience are indirect (through independent dealerships) and occur in less-than-happy circumstances (for example,

when repairs are needed or a vehicle recall is required). Getting involved in service would provide first-hand information of the problems that arise in service and a quick feedback all the way to the design/manufacturing phase for possible pre-emption.

In this paper, we study the benefits of manufacturing-service integration. More specifically, we want to quantify the benefits of information sharing between design, manufacturing and sales units of the OEM and the service centres. The vehicle for our study is the system dynamic models. We develop a cause effect diagram for conceptual model and single product model.

It is known that manufacturing systems are discrete event dynamical systems and so are service networks. Together, they form large discrete event networks. But events in manufacturing system occur at much faster rate than in service networks. The production rates are generally much faster than failure rates. For example, an automobile is manufactured in days but it may not go for servicing for months. Also, while designing IMSN one should be cautious about bullwhip effect in the integrated system which could lead to service system exhibiting oscillatory behavior in its resources: repair crews, spare parts, etc. Our study here using system dynamic models demonstrates the benefits of:

1. Use of service centre feedback in product design to improve both the serviceability as well as reliability.
2. Use of marketing and sales information in reducing bull whip effect in service centre resources.
3. Use of service level or service quality feedback to improve the sales.

In this sense, our study here is a significant contribution to the literature on design – manufacturing - service networks.

2. LITERATURE SURVEY

In scanning the literature on IMSN, one has to search from various angles. From the consumer's angle, there are lots of white papers and reports from Accenture and Mckinsey on the benefits of selling solutions rather than products [3, 6, 9, 10]. It is now established through various surveys that profit margins in service are much higher than in product sales. For certain types of industrial equipment such as heavy machines, aircraft engines, etc service is integral part of product sales proposition. Pricing of services, products and product services bundles is an important issue.

From the economic point of view there have been studies on the importance of manufacturing in service economy [7, 8]. There are two other viewpoints, which are important in practice and implementation, the product life cycle and the business process integration viewpoints. The supply chain limits itself to the point of sale of product, which needs to be overcome by integrating the service phase: physically, informationally and organizationally.

As is known, the product goes through the design, production, customer use, service and finally disassembly, reuse and disposal. Most attention in Product Lifecycle (PLC) is given to the direct chain of events i.e. from the design to the customer but the service and recycling are not integrated into other phases of PLC. In service literature, preventive maintenance [12], the repairman problem [12], spare parts management [1, 11] have received lot of attention. Queueing theory, discrete-event simulation and mathematical optimization techniques are widely used to study these problems. Cohen et al [4] build a product life-cycle model to study a set of strategic choices facing the manufacturer of durable good as he designs the joint product/service bundle. They adopt a competitive (as opposed to single-firm optimization) framework, where there is competition for the provision of after-sales service between the manufacturer and an independent service operator. The product price and service quality/price are characterized by equilibrium to a sequential game. Cohen et al. [5] report on a benchmarking study of 14 companies in electronics, computing and communication industries. They observe that after-sales service revenues constitute 30% of product sales revenues. Optimization of field service operations has received attention in the business literature [2].

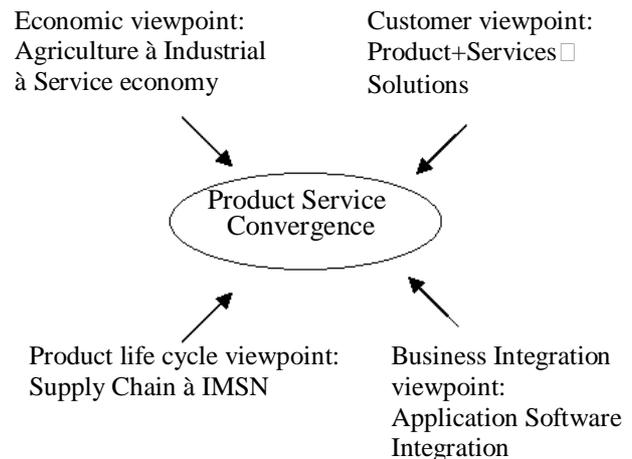


Figure 1: Perspectives on Product Service Convergence

Our attention here is on showing the benefits of sharing information between OEMs and the service providers in terms of improvement in service levels and better service capacity planning.

3. MODEL DESCRIPTION

We consider manufacturing–service network for a complex product with one manufacturing company and several geographically distributed service centres as shown in figure 2.

The manufacturer designs the product, produces it and sells the same to a pool of customers. The product could be automobiles, agriculture equipment like tractors, medical equipment, aircraft engines, ink jet printers, etc.

The common feature is they are all durable products that require after sales support for preventive maintenance, spare parts and also breakdown servicing. Three kinds of maintenance and repair services are possible: self service performed by the customer with on-line help, preventive maintenance where the product performance is monitored by in built diagnostics or by periodic checks, and break down service generally conducted by the service centre. The breakdown service could be either major or minor and the repair time depends on the type of failure. The repair centre has generally the following physical resources: repair crew, disassembly tools, diagnostic equipment, and spare parts. The customer interface of the repair centre is through a call centre or help desk. The informational resources include the data base containing equipment history, types of problems reported for this equipment as well as similar ones installed elsewhere, the knowledge from the crew reports, etc. The data base also may contain information on the spare parts, and can access the CAD designs and the manufacturing data if needed. The resources of the service centre determine the service levels such as the delivery time and the quality of service etc. The service quality as well as the service centre profitability is monitored for possible improvement. The service centre capacity consists of the physical and human resources and this can be changed by hiring or firing the repair men. In this study we assume that the repair centre capacity can be increased with a delay of three months and decreased as and when necessary. Transferring the regional sales information to service centre would help them increase the service capacity in timely manner.

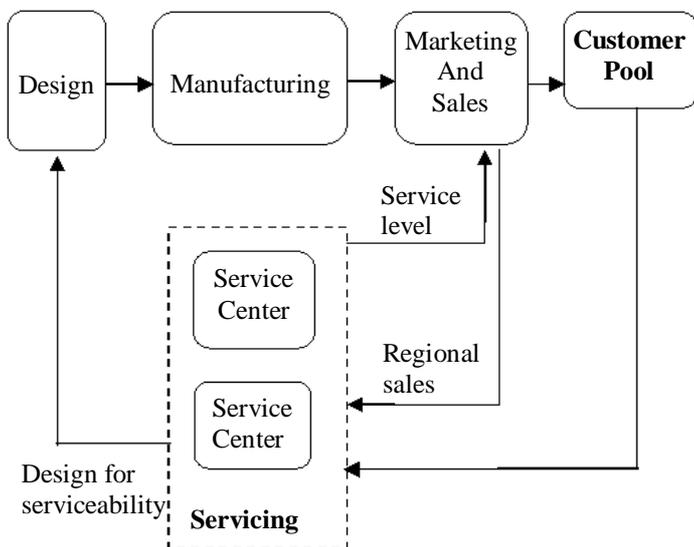


Figure 2: A typical manufacturing-service network

It is known from the literature that after sales service quality influences the product sales particularly from the past customers. We incorporate this in to our system dynamic model by making the service quality as a variable influencing the product sales. Also, the service centre and product design interaction can lead to design

of high reliability products, products with fault diagnostic features, modular products with ease in serviceability etc. We use a utility function to show the positive influence of the service levels on product design which will result in more product sales.

Information exchange between design, service centre and marketing and sales will lead to improvement in the performance of system by bringing coordination. Service centre resource planning will be enhanced by information about regional sales. Similarly, service levels provided by repair centre will improve the sales forecasts of marketing department. We will present three models to demonstrate this.

4. CONCEPTUAL MODEL

We first present a simple model to show benefits of manufacturing, marketing and service centre collaboration. The model demonstrates that sharing information between manufacturing and service phases can have a huge impact on the performance of the system. The simple manufacturer-service network with elements: production, population, sales, service level, repairs, etc. is considered here for illustration. The cause-effect diagram for this is shown in figure 3.

Goods are produced at a constant rate and sales take place according to the utility function. Excess goods produced are inventoried and sold goods enter the population. Whenever a product fails, they come for repairs and service centre makes use of service capacity to perform the repairs. As the population of sold goods increases, more number of products comes for repair and the service capacity needs to be synchronized with population growth to maintain quality of service. Loop 1 in figure 3 shows that as the population increases, there will be more number of products for repair. This would in turn lead to lower service levels and lower the product sales. The service level can be maintained by controlling service capacity in appropriate manner. We explore two mechanisms for controlling service centre capacity in order to maintain the required service level:

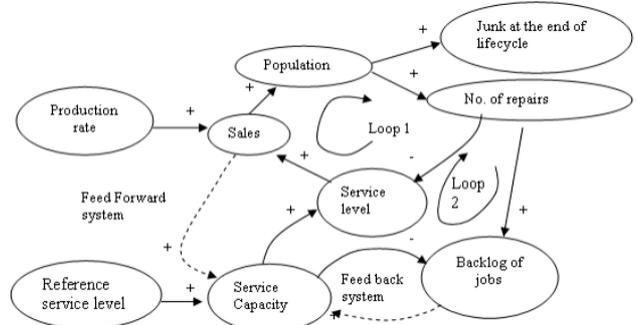


Figure 3: Cause-Effect diagram for Conceptual Model

1. **Feedback system:** Loop 2 is the feedback loop that reduces the backlog of jobs. Whenever jobs get accumulated for repair, the service capacity procurement process is initiated. The new capacity arrives after a fixed

lead time from the time an order is placed. This is the reactive, feedback system and most service centres would typically operate in this fashion when they have no information about product sales. The equation used by the feedback system for capacity requisition is:

$$\text{Capacity ordered on any day} = (\text{Rate of arrival of new jobs}) * 1.1 + (\text{Backlog of jobs}) / 5 - \text{Existing Capacity} - \text{Capacity in the process of procurement}$$

Everyday the service centre is checked for overcapacity and excess capacity is removed according to the following equation:

$$\text{Amount of capacity freed} = \text{Current Capacity} - 1.2 * (\text{Job arrival rate}) - \text{Backlog}$$

2. **Feedforward system:** Sales information can be used to forecast the growth in population and calculate the incremental service capacity required to support the added population. Taking into account the lead-time in procurement of capacity we can ensure timely increase in service capacity. Information visibility that comes with IMSN allows us to have feedforward control to add the service capacity.

We have assumed the lead time to be 120 days and every 120 days we forecast the population growth. The required extra capacity is the population growth multiplied by the failure rate.

$$\text{Capacity ordered every 120 days} = \text{Population growth in 120 days} * \text{Failure rate}$$

Overcapacity is removed according to the following equation:

$$\text{Amount of capacity freed} = \text{Current Capacity} - 1.2 * (\text{Job arrival rate}) - \text{Backlog} - \text{Capacity to support expected population growth}$$

The dotted arrows in figure 3 indicate the feedback and feedforward mechanisms in the cause-effect diagram. We use the cause-effect diagram in figure 3 and Ithink software to simulate the system dynamic model. We assume the following values:

1. A production rate of 1000 units/day is assumed.
2. Car life = 10yrs or 3650 days
3. The service level is the only factor that influences the sales. $\text{Product Sales} = (\text{Mean_Service_level} - 0.7) * 10000 / 3$.
4. Mean number of failure per day = $0.01 * \text{Population}$
5. Acquisition of service capacity takes 120 days because of the delays involved in hiring and training.
6. We assume the product life-cycle to be 10 years long.
7. For performing one repair job, one unit of capacity is engaged for a day.
8. Service level for a particular day is the ratio of number of repairs completed in that day to total number of jobs to be completed (including the backlog jobs). Since the number of repairs completed is at most equal to

the capacity, $\text{Service level} = \text{Min}(\text{Capacity} / \text{Total number of failures}, 1)$.

Under the above assumptions, the ideal service centre capacity = mean number of failures = $0.01 * \text{Population}$
We simulate both feedback and feedforward strategies outlined above.

Feedback system (system 1) responds to accumulation of jobs by increasing the service centre capacity. Since there is a delay in acquiring service capacity and making it operational, in the interim period, the service level declines and jobs accumulate. To clear this up quickly, excess capacity needs to be hired. Thus the reactive feedback system creates a bull whip effect: a sudden increase in product sales will after a lag create a need for excess service capacity. Because of the set up times involved in increasing the service centre capacity, the service levels decline bringing down the sales. This is clear from figure 4. The kind of fluctuations in the service capacity shown in the figure can be prevented by collaboration between the manufacturing and the service centre.

Figure 5 compares the two systems: 1) System with only feedback. 2) System with both feed forward and feedback. The feed-forward system is able to maintain very high service levels while the feedback system shows fluctuations. This is because feed-forward system accurately predicts the required capacity and adds it taking into account the lead time.

On the other hand, feedback system shows considerable variation in the service level, which in turn effects the population growth. The graph in figure 4 illustrates the significant benefits. We observe that because of the high service level, the population in system 2 has more sales than system 1, which is reflected in the populations of the corresponding systems.

An IMSN based on a feedback mechanism responds to the accumulation of jobs by increasing the service center capacity. However since there is an inherent delay in acquiring and deploying service capacity, the impact of service capacity addition is not immediate. Consequentially the service levels continue to decline and jobs continue to accumulate in the interim period. Hence, to rapidly clear the backlog of jobs and to bring the service up to the desired levels, sometimes more capacity than required needs to be added. Thus the reactive feedback system results in a bull whip effect: *a sudden increase in product sales after a lag creates a need for excess service capacity*. Because of the leadtimes involved in service capacity addition, the service centres are unable to maintain their service levels and accordingly the sales decline as shown in Figure 4. The fluctuations in service capacity observed in the system (as shown in Figure 4) can be prevented by promoting collaboration between the manufacturing, sales and the service operations of the IMSN.

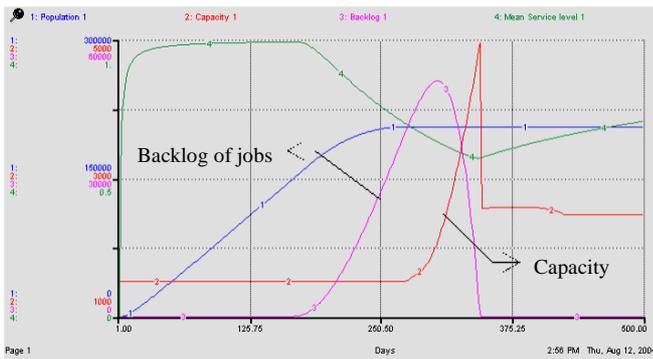


Figure 4: Buildup in service capacity to clear accumulated jobs

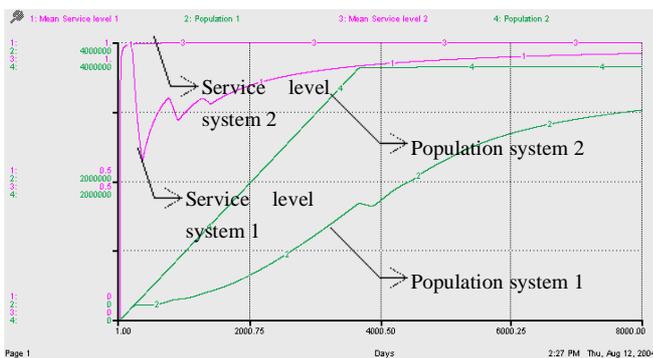


Figure 5: System with Feed-forward and Feedback maintain higher service level than system with only feedback

5. MODELING IMSNS WITH A SINGLE PRODUCT

The basic model presented in Section 4 is simplistic in its assumption that the consumer's decision to buy a product is based primarily on the quality of the service. However, in reality the purchase decision is influenced by multiple factors such as the price and quality of both the product and the service and the quality and length of the warranty period. In this section we attempt to incorporate these features into a more realistic model of an IMSN that manufactures and services a single product.

For our analysis we assume that the utility of a product to a consumer is a linear function of product price and quality, service price and quality and the length of the warranty period.

$$\text{Utility} = -\text{Product price} + 1000 * \text{Product quality} - \text{Service price} + 1000 * \text{Service level} + \text{Free service contract}.$$

Product price and service price have a negative influence on the utility function since higher prices deter customers from buying the product. Similarly, free service contract periods, better product and service qualities, have a positive influence on the utility as perceived by customers. Furthermore, we also incorporate a reference utility function, similar to the utility described above, which defines a target utility level for the company to aspire towards. In reality this reference utility is

equivalent to the utility of the products offered by the competitors. In this manner we are able to account for competitive pressures facing an IMSN. The reference utility level also provides a goal for the company to work towards, by improving the product and service value proposition. As can be expected, the better the product offered by the IMSN, relative to the competition, higher its sales. Hence, the number of sales that take place in a day is directly proportional to the difference between the reference utility (utility of competitive offerings) and utility of the product offered by the IMSN. If a product service bundle offers superior value the demand for the product will expectedly increase. However increased sales and growth in the product population will over time result in more service demands putting a strain on service levels and quality, requiring some kind of monitoring process to remedy the service deterioration. The influence diagram for an IMSN manufacturing and servicing a single product under the above operating assumptions is presented in Figure 9.

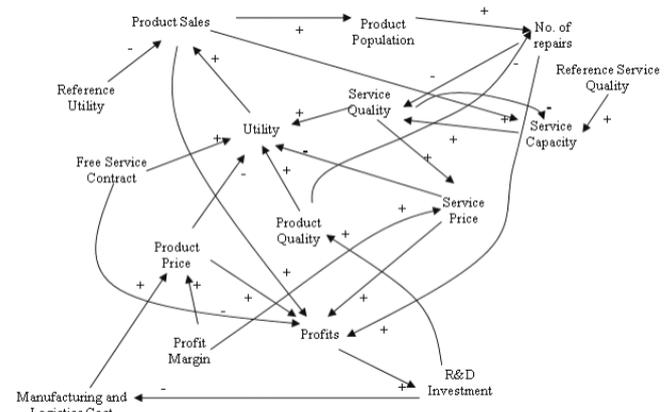


Figure 9: Influence diagram for IMSN with single product

The profits of the IMSN are derived from the sales of both products and services, less the cost of production, service provision and cost of providing free services during warranty. In addition, we assume that a fixed percentage of the profits are continually invested in R&D. These investments bear fruit at a later time in the form of improved product quality, reliability and reduced product prices, in turn stimulating further sales and profits. Concurrently, as the product quality improves, the reliability of the product increases and the number of failure events and consequentially repairs is reduced. In this manner, the concept of design for serviceability is included in our model of an IMSN. However for the sake of simplicity, we have assumed aggregated models for service centres, each of which are geographically distributed. One key differentiating factor in the extended model presented in this section, as compared to our earlier analysis, is the assumption that the product failure rate increases with the age of the product. With an increase in product sales over time the population grows at a higher rate, resulting in greater number of repair jobs. Also, as a product gets old it comes more often for repair. Hence, to predict the load on the service centre we need

to track both product sales and the age of the installed base.

Additionally, we incorporate the feedforward and feedback mechanisms, explained in the earlier sections, for controlling the interaction between the manufacturing and service sectors and for ensuring that service levels are maintained at the required level. For example, in a feedback based system, if the service level drops below the reference service level, service capacity is added. On the other hand, in a feedforward system, the sales data is employed to forecast and plan for additional capacity requirement.

The above model was simulated in IThink to demonstrate that the feed-forward system is better than the feedback system in terms of its ability to improve service levels and the bottom-line for the IMSN. For the simulation study, it was assumed that the lead time for service capacity addition was 120 days and the initial capacity was set to be adequate to meet all demands occurring in the first 120 days of operations. Based on growth of the population and age of installed base a forecast is made for number of failures per day after 240 days and the required extra capacity is added as appropriate. From the influence diagram it may be noticed that the dynamics within this extended and realistic model are more intricate than the basic model presented in Section 4 (Figure 3), establishing the fact that it is harder to predict the behavior of realistic IMSN models. Our analysis was based on the following parameters and assumptions.

1. One unit of service capacity costs 100 units of money per period.
2. The price charged for service dependent on the service level offered as dictated by the following relationship.

$$\text{Service Price} = \text{Service capacity cost} * (1 + \text{service level}).$$
3. Profit margin of 10% incorporated in the price of products
4. Product quality & Manufacturing cost improves non-linearly for incremental investments in R&D, meaning that the ROI for R&D investments is more than 100%.
5. The probability of a product failing increases over time.
6. The product becomes obsolete in 6 years, in a non-linear manner, as it rapidly becomes outdated in its first few days before being gradually replaced over time.

The service levels for IMSNs with only feed-back and with both feed-back and feedforward were simulated and compared as shown in Figure 10. Clearly an IMSN with feedforward and feedback mechanisms is able to provide consistent service levels, inspite of the dynamics involved in the system. It may be observed from Figure 11 that the service capacity growth in an IMSN with a feedforward system is more planned in comparison to an IMSN with just a feedback system. Since one unit of service capacity is needed to complete a single repair job

in a day, the service capacity over time in a well synchronized manufacturing service network, such as one with a feedforward system, will trace the number of incoming repairs jobs. However, due to the long lead-times in service capacity acquisition the service capacity increases only in a stepwise manner. It may also be noticed that the excess service capacity added in an IMSN with only feedback is eventually eliminated as the backlog of service jobs is cleared. The models for IMSNs with both feedback and feedforward were simulated over the entire lifecycle of the product and their profits compared. The profits in an IMSN with both feedback and feedforward mechanisms are 50% higher than the profits for an IMSN with only feedback.

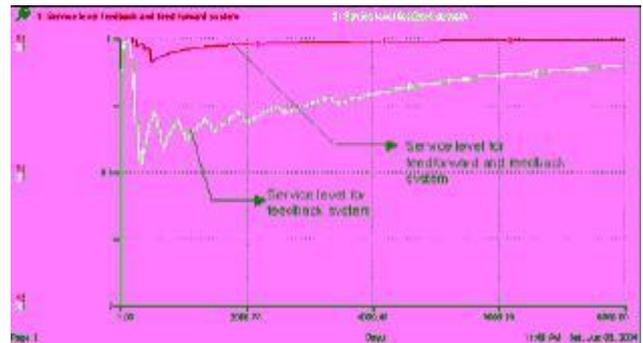


Figure 10: Comparison of service levels for an IMSN with feedback and one with both feedforward and feedback.

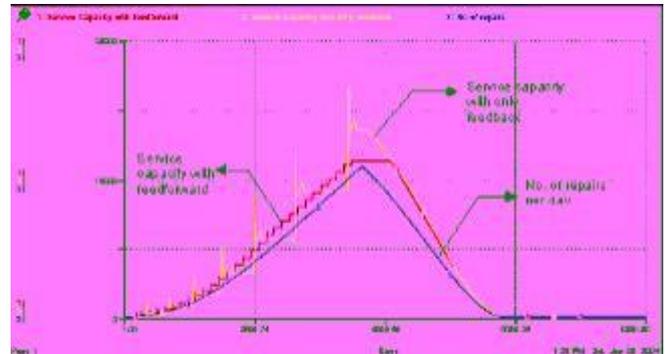


Figure 11: Comparison of service capacity for system with only feedback and system with feedback and feedforward

6. CONCLUSION

In this paper, we have studied the information sharing between manufacturing plant and service centres. While there are several studies in the form of business cases and white papers on benefits, ours is the first systematic study providing a framework for the analysis. In particular we study the beneficial effects of service centres feedback to design teams and sales and marketing departments. Feed forwarding the sales information to the service centres to reduce the bullwhip effect on the repair crews and spare parts.

In terms of the future, we develop multiple product

models and determine optimal time for the new product introduction. We also hope to develop queueing network and Petri net models to further study and develop analytical models.

REFERENCES

- [1] Aberdeen Group, "Service Parts Management: Unlocking the value and profits in the service chain", September, 2003.
- [2] Aberdeen Group, "The Field Service Optimization Benchmark Report", June, 2004.
- [3] Bundschuh, Russel G., Dezvane, Theodore M., "How to make after-sales services pay off", The Mckinsey quarterly, 2003, No. 4.
- [4] Cohen, M., Whang S., "Competing in product and service: A product life-cycle model", Management Science, Vol. 43, No. 4, April 1997.
- [5] Cohen, M., Y Zheng, and V Agarwal, "Service Parts Logistics: A benchmark analysis", IIE Transactions, Special issue on supply chain management, 1997.
- [6] Dennis, Michael J., "Service Management: Buidling profits after the sale", Supply Chain Management Review, January 2003
- [7] Jane Marceau, Cristina Martinez, "Selling Solutions: Product-service packages as links between new and old economies", DRUID Summer Conference on "Industrial Dynamics of the New and old Economy – who is embracing whom?", Copenhagen/Elsinore 6-8 June 2002
- [8] John Zysman, "Production in the Digital Era: Commodity or Strategic Weapon?", Berkeley Roundtable on the International Economy, 2002.
- [9] Juliet E. Johansson, Chandru Krishnamurthy, and Henry E. Schlissberg, "Solving the solutions problem", Mckinsey quarterly, 2003, No. 3.
- [10] Knecht, Thomas, Leszinski Ralf, and Weber, Felix A., "Making profits after sale", The Mckinsey quarterly, 1993, No. 4.
- [11] Servigistics, "Service Parts Management: The untapped opportunity", November, 2003.
- U Dinesh Kumar, John Crocker, J. Knezevic, M El-Haram, "Reliability, maintenance and logistic support –A life cycle approach", Kluwer Academic Publishers, 2000

A Study of Corporate Managerial Module and Information System Orienting Agile Virtual Enterprises

Tong Shu¹, Shou Chen²

College of Business Administration, Hunan University, Changsha, Hunan 410082, China

¹ shutong@hnu.cn, ² chenshou@hnu.cn

ABSTRACT

The operation modules and information system construction are essential contents in agile virtual enterprises. Based on the analyses of the characteristics, organization forms and operation processes in agile virtual enterprises, studies are extended to the corresponding information systems, and the development of systematic logic modules and kernel methodology is also proposed in the paper.

Key Words: Agile virtual enterprises, Operation modules, Management information systems

1. INTRODUCTION

With the development of information technology, there are great changes in global economic environments. Consumers prefer the concept of individuality and diversity, and the uncertainty of the market demand is increasing greatly. All these call for the enterprises to have flexibility and adaptability in the market. The changeable markets also call for the enterprises to have simple managerial structures, which are easy to reorganize and adjust. The customers' needs of individuality and diversity call for the enterprises to have creativity to develop new products. Consequently, the enterprises and management review the traditional managerial modes and competition models.

In 1991, supported by American National Defense Department and 13 well-known enterprises, Roger and Rick Dove, together with more than 100 experts in Iacocca Institute of Lehigh University in USA, proposed a report entitled Manufacturing Strategies of the 21st Century before the American Congress. A new manufacturing model was put forward in the report, i.e. Agile manufacturing on the basis of agile virtual enterprises (VE). The fundamental concepts of VE are that the enterprises make good use of the latest achievement in information technology and social resources and strategically join the competitive advantages of enterprises, concentrating on the development of new products and better management. The management, development, manufacturing and marketing of products are implemented by means of VE, which vary with the markets and become the kernel part of agile manufacturing.

2. CHARACTERISTICS OF VE

The motivation of the emergence and development of VE is to fulfill the joint destinations and to achieve the agile effects. The characteristics are integration, agility, virtuality and time-validity.

ð *Integration of VE:* Integration indicates that the relevant enterprises with varied managerial advantages dedicate their competitive privileges to the VE, conforming and exerting their unity effects in order to develop prominent competitive edge. This can be realized by means of resources conformity and resources complementation. Integration is the primary function of VE and also the major approach to the system reconstruction.

ð *Agility of VE:* Agility indicates that the enterprises are able to react and accommodate to the demands of the dynamic markets by grasping the market opportunities and promoting creativity, such as being reconfigurable, being reusable, and being scalable, namely RRS characteristics. From the perspective of business management, these characteristics refer to connection, cross-organizational participation, manufacturing flexibility and managerial relevance.

ð *Virtuality of VE:* In order to fit the dynamic markets, the structure between members is not stable, but they can reconstruct in terms of changing objectives, replacing products and varying environment, namely organizational virtuality. The VE have functions such as manufacturing, assembling, management and finance, but there are no corresponding internal organizations, namely functional virtuality.

ð *Time-validity of VE:* VE are based on the market opportunities, uniting or dismantling in terms of business opportunity. Their members are autonomous and independent.

3. ORGANIZATION OF VE

The managerial approaches of VE focus on the production procedures, adopting the forms of teamwork. There are four tiers, namely Virtual Enterprise (VE), Virtual Group (VG), project team based on products or services (TEAM) and Basic Organization Unit (BOU).

The first tier is the virtual enterprise made up of numerous functional projects. The second tier is the cooperation between enterprises made up of many functional projects by information network, and their methods of cooperation are virtual cooperation, interrupted compatibility, transference and supply chains. The third tier is the mobile organization between

enterprises by establishing basic organization units. The fourth tier is the basic manufacturing organization units made up of equipment, staff and technology. All the four tiers represent the many-to-many mapping between the physical material and dynamic associations within the enterprises, illustrated as Figure 1.

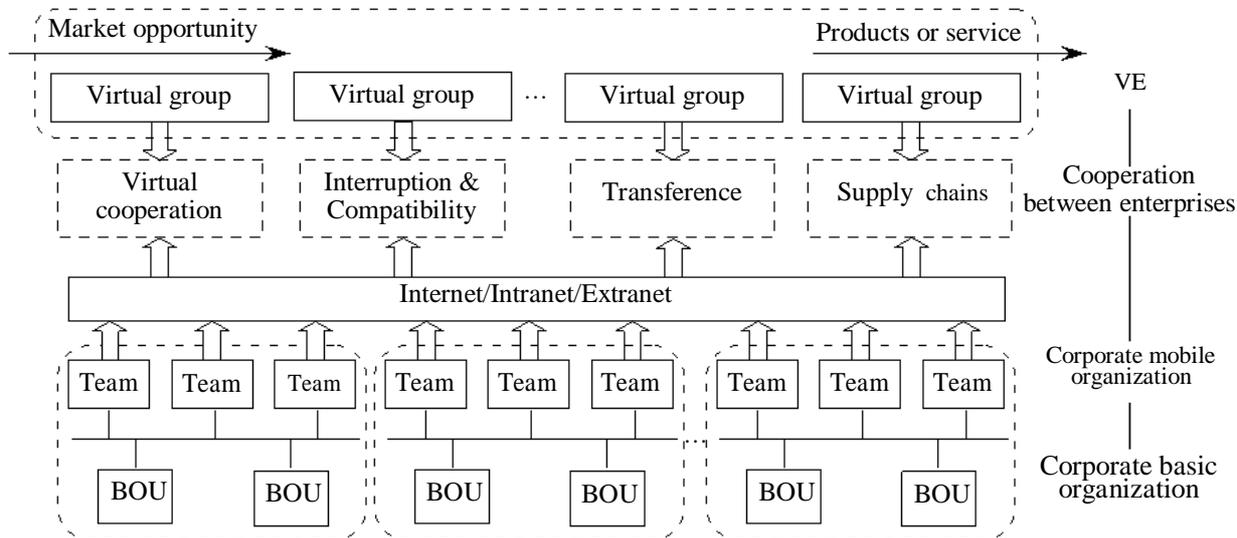


Fig.1 The Description of Tier Structures of VE

In the construction of VE, the procedure is as follows: the description of the business opportunity → definition of the business processes → the confirmation of the essential resources → selection of the business partners → decision on the cooperation strategies → the allocation of the profits and risks → management of dynamic organization. The whole procedure involves priority combination issues between internal organizations, between internal and external organizations, and between members among other cross-disciplines, etc. The ultimate goal is to establish the effective cooperation by setting up network system.

4. MANAGEMENT PROCESSES IN VE

The management of VE is proposed by one or several enterprises that manipulate the essential resources (e.g. vital technology, markets or brands), and members provide their predominant resources and share the risks and profits. Consequently, the management of VE can be decomposed into 5 phases.

ð *Identification of Business Opportunity:* It is the starting point to grasp and identify the business opportunities. There are varied approaches for VE to acknowledge the business objects, namely market investigation and prediction, analyses of the market feedback, research and development of new products, analyses of market competition and social environments. The VE will discover the potential products and service in the market in

terms of the product and technology development strategies of their members.

ð *Structural Analyses of VE:* The advantages of Virtual Enterprise are gained by means of sharing information between enterprises and harmonizing the production and management processes within enterprises. The member enterprises achieve their supply of chance products or services by their own kernel capability. VE ensure the demand resources in light of the characteristics of the chance products, involving the resource distribution and the conditions of Internet communications between resource possessors. The coalescence capability plays a vital role in the success of VE.

ð *Selection of Partners and Establishment of VE:* VE make profits through time-limited cooperation and reset of kernel resources, and they also concentrate the market objectives, investigate and evaluate the relevant potential partners. In particular, the evaluation of the vital capabilities is the basis for the selection of partners. VE observe such principles as resource complementation, profit sharing and risk partaking, define the allocation forms, ensure the relationship between member enterprises and accomplish their establishment. Sound allocation programs of risks and profits are the important bases for the establishment and normal operation of VE.

ð *Managing, Tracing and Recording:* Members of VE have cooperation and competition. In terms of

regulations and market situations, VE adjust organizational structures dynamically. If the feedback suggests that there are great changes in the member enterprises and markets, VE will reorganize and even disorganize.

- ò *Reckoning and Disorganizing*: When VE complete the specific products and services of the business opportunity, they will reckon the correspondent expenditure, rights and interests in light of the contracted interest-allocating program, involving the apportionment of management expenditure, development expenditure, and manufacturing expenditure. Additionally, VE reckon the funds in the business management, and tackle with the belonging of the new technology and its future application. These processes will occur during the operation or after the disorganization of VE.

5. MIS OF VE

In the relatively balanced market environment, the traditional information system plays a vital part in the business management of enterprises. However, in the chaotic changeable environments, particularly in the virtual manufacturing module that is the major structural form in VE, the business information systems are required to accommodate quickly to the changes and development of enterprises. Compared with the traditional information system, the prominent characteristics of the management information system (MIS) of VE are the complete distribution of the system information, structures and functions, dynamic reorganization and agility.

5.1 Characteristic of MIS of VE

- ò *Openness and Compatibility*: VE is a new type of virtual business that takes the advantage of the network information technology and breaks the regional barriers. Information flows between member enterprises and within enterprises, which requires that the Management Information System within VE is not blocked and that diversified management information system can be integrated.
- ò *Flexibility*: The reconstruction of dynamic business operation and the reorganization of the units require that the corresponding MIS is rather flexible for VE so that it can execute functional integration and adjustment, and meet the needs of new business flows after the dynamic reorganization.
- ò *Robustness*: MIS in Virtual Enterprise is quite sophisticated, which meets the demand of the information integration and management between and within member enterprises. It should be reliable and effective, and it is robust in supporting large-scale cooperation on the basis of multi-disciplines and multi-tiers.

5.2 Logic Module of MIS in VE

In the logic module of MIS in VE, there is information flow within enterprises, between member enterprises, between enterprises and the environment, between VE and the environments, illustrated in Figure 2.

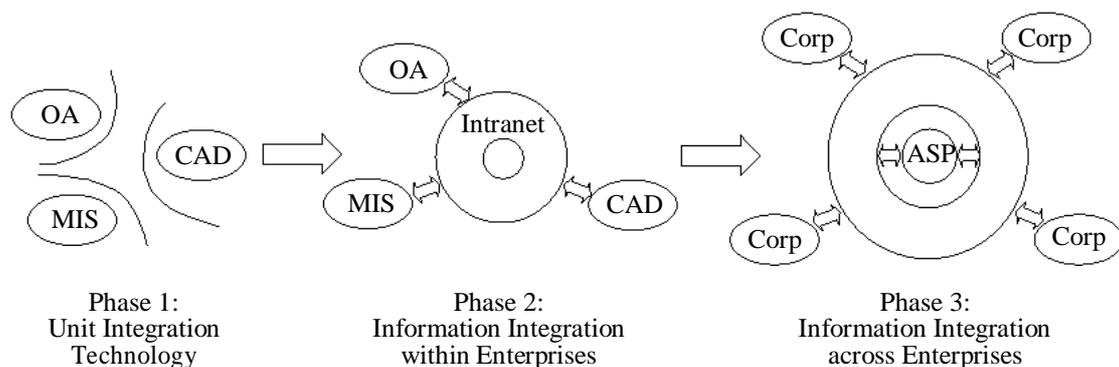


Fig.2 Logic Module of MIS in VE

Information flow within enterprises is based on Intranet. Their MIS is responsible for the collection and settlement of the basic data of enterprises, corresponding to the treatment of the businesses and interior information in all the business management departments and all the production phases. It provides the enterprises with environment for information sharing, communication, cooperation and decision-making. Object-Oriented Methodology has such characteristics as encapsulation, inheritance,

multi-states, modifiability, dynamic connection and being reusable, so that it is convenient to develop the renovative and connective MIS. The designing concept is based on the business opportunity and production missions of the enterprises. Objects are selected from the object-bank, orderly business processes and sound business organizations are matched. The parameters are configured corresponding to subsystems such as people, property, materials, production, supply, sales,

prediction and decision-making of business management.

Information flow between member enterprises is based on Extranet. Their MIS is outstretched, and it is a cooperative network that makes enterprises connect and accomplish their mutual goals. It can be a bridge between public Internet and specialized Intranet. It assumes the information communications between intimate member enterprises. It can also be viewed as a part of Intranet between the enterprise and its cooperative enterprises.

Information flow between VE and the environment is based on Internet. Their MIS is applied to information promulgation and communication, advertising systems, e-commerce systems, which can be fulfilled through setting up corresponding external server of the enterprises. Regarding the security of the system, data bank safety of MIS and network system safety are the two major factors to be considered.

5.3 Core Methodology of MIS in VE

This sort of information flow is integrated in enterprises on the basis of Internet/Intranet/Extranet technology, and supported by significant technology, namely object-oriented methodology (OOM), standard for the exchange of product model data (STEP), a module-setting language (EXPRESS), common object request broker architecture (CORBA).

OOM expresses through objects. It is corresponding to the characteristics of VE such as versatile information data and enormous data quantity. Also, it improves the encoding efficiency, system maintenance, and the capability to be reusable.

STEP is to offer data module in establishing distributive information system for VE, and then it is transferred to the traditional package treatment data through corresponding formats.

EXPRESS is to present objects, information units of the objects, and the operation that confines and permits the objects. It is a formal information module-setting language, and it provides the datum information that has the engineering datum structure and distributive information system conforming to STEP.

CORBA is an essential communication mechanism that can operate between distributive objects, and it is a canonical system construction that can make an individual visit to objects possible in a separate structural systems. It defines a structure for object request broker (ORB), and it meets the demand of inter-operation between management objects and applications, and the compatibility of the separate structural systems. Additionally, it can maintain the relative independence between member enterprises.

Information Treatment Methodology of agile virtual enterprises also involves:

- ò open inter-operable underlying network protocols (TCP/IP, HTTP, RMI, SOAP),
- ò open distributed object oriented middleware services (J2EE Framework, ActiveX Framework),
- ò information/object exchange mechanisms and tools (XML, ebXML, WSDL),
- ò standardized modeling of business components, processes and objects (EJBs, OAG and OMGs Business Objects and Components),
- ò business process modeling tools and languages (UML, UEMML, WfMC XML-based Business Language, PSL),
- ò open and standard business process automation and workflow management systems (WfMC, OMG-JointFlow, XML-WfMC standards, many commercial products),
- ò standard interfacing to federated multi-databases (ODBC, JDBC),
- ò intelligent mobile agents (FIPA, OMG-MASIF, Mobile Objects),
- ò open and standard distributed messaging middleware systems (JMS, MS-Message Server, MQSeries, FIPA-ACC),
- ò XML-based E-commerce protocols (BizTalk, CBL, OASIS, ICE, RosettaNET, OBI, WIDL),
- ò web integration technologies (Servlets, JSP, MSASP, XSL).

The current security mechanisms can be fulfilled through safety methodologies such as the state and behavior of datum bank, rich semantic association, the specification of managerial purview, datum encryption methodology, firewall technique, network security scanning methodology, network invasion testing methodology and hacker inducement methodology.

6. CONCLUSION

The traditional management theory concentrates on the internal relationship in the enterprises, whereas VE focus on reshaping the relationship between enterprises. The managerial methodology of VE replaces sequential management with juxtaposition managerial procedures, and replaces manipulating management with concentrated and cooperative managerial methodology. VE are the advanced organizational management behavior and market competitive strategies directed by the concept of agility, accommodating to competition and manufacturing development. The soul and kernel of VE are agility and integration.

The development and application of the Internet technology make it possible for the former enterprises to transfer the specialized information from the comparatively higher tiers to the lower tiers, and to feed back the information from the mid or lower tiers of the enterprises to the high-tier management through excavating data. Virtual enterprise modules enable

enterprises to reconstruct their production, supply, sales and service flow. All these demand that enterprises completely construct their MIS to support business management of VE. Enterprises are also required to build up concepts of competition and competition, and to integrate the external advantages through taking the advantage of internal resources, and to react quickly to the market demands, and thus to enhance the competitive edge of the kernel enterprises.

REFERENCES

- [1] Luis M. Camarinha-Matosa, Hamideh Afsarmanesh, "Elements of a base VE infrastructure", *Computers in Industry* Vol. 51, No. 3, pp139-163, 2003.
- [2] Wang, Qing, Yung, Kai Leung, Ip, Wai Hung, "A hierarchical multi-view modeling for Networked Joint Manufacturing System", *Computers in Industry*, Vol. 53, No. 1, pp59-73, 2004
- [3] Mangesh P. Bhandarkar, Rakesh Nagi, "STEP-based feature extraction from STEP geometry for Agile Manufacturing", *Computers in Industry*, Vol. 41, No. 1, pp3-24, 2000.
- [4] Srinivas Talluril, R.C. Baker, Joseph Sarkis, "A framework for designing efficient value chain networks", *Production Economics*, Vol. 62, No. 5, pp133-144, 1999.
- [5] Olivier Perrin, Claude Godart, "A model to support collaborative work in virtual enterprises", *Data & Knowledge Engineering*, Vol. 50, No. 7, pp63-86, 2004.
- [6] H.C.W. Lau, K.S. Chin, K.F. Pun, A. Ning, "Decision supporting functionality in a virtual enterprise network", *Expert Systems with Applications*, Vol. 19, No. 3, pp261-270, 2000.
- [7] Ronald C. Beckett, "Determining the anatomy of business systems for a virtual enterprise", *Computers in Industry*, Vol. 51, No. 3, pp127-138, 2003.
- [8] Troy J. Strader, Fu-Ren Lin, Michael J. Shaw, "Information infrastructure for electronic virtual organization management", *Decision Support Systems*, Vol. 23, No. 5, pp75-94, 1998.
- [9] George L. Kovacs, Paolo Paganelli, "A planning and management infrastructure for large, complex, distributed projects—beyond ERP and SCM", *Computers in Industry*, Vol. 51, No. 3, pp165-183, 2003.
- [10] Xu Xiaofei, "Organization Methodology System of Agile Virtual Enterprises", *Computer Integration Manufacturing System—CIMS*, Vol. 5, No. 1, pp7-12, 1999.
- [11] Wu Naiqi, Sun Jian, "Mission Allocation among Members in Agile Virtual Enterprise Reconstruction", *Computer Integration Manufacturing System—CIMS*, Vol. 7, No. 8, pp1-5, 2001.
- [12] Zhang Changhao, Xia Anbang, Zhou Hangbing, "A Study of Enterprise Information System in Agile Virtual Enterprises", *Computer Engineering and Application*, No. 8, pp211-214, 2003.
- [13] Zhan Dechen, Ye Dan, Xu Xiaofei, Li Quanlong, "Enterprise Module of Agile Virtual Enterprises", *Computer Integration Manufacturing System—CIMS*, Vol. 5, No. 3 pp11-15, 1999.
- [14] Zhou Hui, Qin Zheng, Liang Wei, "A Study of System Structure of Distributive Information System in Agile Virtual Enterprises", *Transactions of Northwest Technical University*, Vol. 23, No. 1, pp29-34, 2003.
- [15] Zhang Shensheng, Gao Guojun, "Virtual Enterprises and Agile Supply Chains", *Computer Integration Manufacturing System—CIMS*, Vol. 5, No. 2, pp1-5, 1999.
- [16] Shi Chunsheng, Li Xiangyang, Fang Shufen, "Organizational Format and System Design of Agile Virtual Enterprises", *Transactions of Managerial Science*, Vol. 3, No. 2, pp21-26, 2000.
- [17] He Yibin, Ma Dengzhe, "A Study of Prototype System of Production-oriented Agile Virtual Enterprises", *Machine Industrial Automation*, Vol. 20, No. 5, pp13-15, 1998.
- [18] Han Jian, Wu Cheng, Fan Yushun, "Information Basis Constructions Supporting Agile Virtual Enterprises", *Transactions of Qinghua University*, Vol. 38, No. 10, pp61-64, 1998.

A Study of Single-vendor and Multiple-retailers Pricing-Ordering Strategy under Group-Buying Online Auction

J. Chen¹, Yunhui Liu¹, Xiping Song²

¹ School of Economics and Management, Tsinghua University, Beijing 100084, China

² Department of System Engineering and Engineering Management
The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China
{chenj, Liuyh3}@em.tsinghua.edu.cn, xpsong@se.cuhk.edu.hk

ABSTRACT

The supplier and buyers, with different objectives and self-interest, are separate economic entities acting independently and opportunistically to maximize their individual profits. In this paper, a GBA model in the B2B market is studied, where one supplier faces 2 different retailers, who cooperate in the order decision making. Firstly, the optimal ordering decision of the retailers was analyzed. Then, from the perspective of the supplier, the optimal pricing strategy of the supplier is also studied. Finally, it is concluded that the group buying online auction is a useful and efficient pricing mechanism in the B2B e-market, under which, all members of the supply chain will improve their payoffs.

Keywords: Group-buying online auction, supply chain, newsvendor model, B2B market

1. INTRODUCTION

The supplier and buyers, with different objectives and self-interest, are separate economic entities acting independently and opportunistically to maximize (minimize) their individual profits (costs). An effective supply chain network requires a cooperative relationship between the vendor and the buyer. Based on mutual trust, the cooperation includes the sharing of information, resources and profit. One of the most common strategies is to setup a pricing policy acceptable to both the vendor and the buyer. (Yang, 2004).

The price discount policy is a subject of significant interest. Starting from Crowther (1964), Monahan (1984) was one of the early authors who analyzed a vendor-oriented optimal quantity discount policy that maximized the vendor's gain; it is done at no additional cost to the buyer. Lee and Rosenblatt (1985) generalized Monahan's model and developed an algorithm to solve the vendor's ordering and discount-pricing policy. Buscher and Lindner (2004) further extend Lee and Rosenblatt's model.

However, their objective function is restricted to maximizing the supplier's profits only. Zahir and Sarker (1991) consider a price-dependent demand function for multiple regional wholesalers who are served by a single manufacturer. Meantime, in recent years, some studies have paid the attention on the perishable product with the policy of price discount. Wee (1999) introduces a deterministic inventory model with quantity discount, pricing and partial backordering when the product in stock deteriorates with time. Papachristos and Skouri (2003) generalize the work of Wee and consider a model where the demand rate is a convex decreasing function of the selling price and the backlogging rate is a time-dependent function, which ensures that the rate of backlogged demand increases as the waiting time to the

following replenishment point decreases.

In our paper, we also analyze the two-echelon supply chain consisting of one vendor and multiple different buyers, in which, the products is perishable. The main difference between our work and theirs is that a particular pricing mechanism, the Group buying online auction, is employed in our study.

Group-buying auction is also a kind of quantity discount form. However, the GBA, unlike the traditional discount, does not lead to price discrimination among different buyers and every buyer will be charged the same price. Each buyer in the GBA will only buy a piece of product which means the price declines monotonically in total purchase quantities of all buyers, and not just an individual buyer's purchase quantities. These are much different from the traditional quantity discount policy.

Moreover, the group-buying auction (GBA), a homogeneous multi-unit auction, has many users on sites such as LetsBuyIt.com and Yabuy.com. Different from traditional auctions, where bidders compete against one another to be the "winner" with the highest price, the GBA enables bidders to aggregate to offer a lower price at which they all "win" (Horn et al. 2000). When using the GBA as the pricing mechanism, it is helpful to refer to traditional auction models. By considering the seller's expected revenue, Vickrey (1961) proves the well-known Revenue Equivalent Principle and Myerson (1979) proves the Revelation Principle. The details about traditional auction studies can be found in Klemperer's (1999) survey. However, the traditional auction theory does not consider inventory because it is not an important issue for those products. Internet expands online auction markets, where almost all products can be auctioned. Hence, our study tries to bridge the inventory management and auction process.

Although the GBA is widely used in the E-market, it lacks a theoretical framework and the study on the GBA is rare. Chen et al. (2002) build a dynamic game model for the GBA, based on which they study the bidders' optimal strategy. They prove that the mechanism is incentive compatible for bidders under the IPV (independent private values) assumption. Kauffman and Wang (2002a) conduct an experimental study of the GBA. They analyze the changes in the number of bids for Mobshop-listed products over various periods. Based on the empirical study of twelve GBA websites, Kauffman and Wang (2002b) point out that because the GBA is still a new concept, the websites must educate consumers to achieve a critical mass. Anand and Aron (2003) devise an analytical model to study the GBA. They compare the posted pricing mechanism with the GBA in different scenarios, e.g., demand uncertainty and scale economies of production. On the other hand, the inventory literature has paid little notice to the marketing trends (Arcelus et al., 2003). Chen et al. (2004b) consider a GBA model for a monopolistic manufacturer selling novel products in the uncertain market and link this new pricing mechanism to the newsboy model.

They, however, all do not study the GBA mechanism in the B2B market, which is a more prosper and profitable segment. Kauffman and Wang (2002b) suggest that the GBA websites oriented toward the B2B market should be better positioned for future growth. In the business practice, Mobshop changed its strategic direction from the B2C to the B2B market (Clark 2001) which is more profitable. According to Anand and Aron (2003), in the B2B market, the GBA has been employed to sell diverse goods and services, such as, furniture, commercial print, computer hardware, software, telecommunications & connectivity, electricity and gas, computer equipment, office supplies and vehicles, vegetables and company travel. To the best of our knowledge, our paper is the first in academia to study the GBA in the B2B e-market. Our effort is to regard the GBA as a useful and resultful coordination strategy in a two echelon supply chain in order to make a more effective and efficient decision in the B2B e-Market environment.

The rest of this paper is organized as follows. In section 2, we build the model with one supplier and 2 retailers. Section 3 deduces the optimal order strategy for the retailers. In Section 4, we study the supplier's optimal discount strategy. In section 5, we summarize the paper and describe the future research directions.

2. MODEL DESCRIPTION

Suppose two retailers, which will cooperate during the ordering decision process, order products from the same supplier and each retailer faces independent markets. First, the supplier will set the price curve (p_1, p_2, l) , where $p_1 > p_2$. Then the retailers will bid their order

quantity. If the sum of the order quantity is equal to or greater than l , then all the retailers pay unit price p_2 to get their order. Otherwise, the unit price is p_1 . Suppose retailer i faces the i th markets with unit sale price S_i , and the demand is a random variable with PDF $F_i(\cdot)$. At the end of the auction, the unsold products are perished with salvage 0. All information is common knowledge and the retailers make their decision simultaneously. i.e., they face a statistic game.

3. THE ORDER STRATEGY FOR THE RETAILERS

When no quantity discount is offered, the supplier's unit selling price is p . Then each retailer faces the situation as the newsboy model and the payoff for retailer i is

$$\int_0^{n_i} S_i x dF_i(x) + \int_{n_i}^{+\infty} S_i n_i dF_i(x) - n_i p$$

The optimal order quantity is $n_i^* = F_i^{-1}(\frac{S_i - p}{S_i})$ and the payoff is

$$\int_0^{n_i^*} S_i x dF_i(x)$$

When using the GBA, the wholesale price is related to the sum of the retailers' order quantity, i.e., the wholesale price

$$p = H(n_1 + n_2 - l)p_2 + (1 - H(n_1 + n_2 - l))p_1,$$

where
$$H(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

Hence the payoff for retailer i is

$$\int_0^{n_i} S_i x dF_i(x) + \int_{n_i}^{+\infty} S_i n_i dF_i(x) - n_i (H(n_1 + n_2 - l)p_2 + (1 - H(n_1 + n_2 - l))p_1)$$

When the two retailers cooperate, their optimal strategy is to maximize the whole profit of both retailers, i.e.,

$$\begin{aligned} \max_{n_1, n_2} & (\int_0^{n_1} S_1 x dF_1(x) + \int_{n_1}^{+\infty} S_1 n_1 dF_1(x) + \int_0^{n_2} S_2 x dF_2(x) \\ & + \int_{n_2}^{+\infty} S_2 n_2 dF_2(x)) - (n_1 + n_2)(H(n_1 + n_2 - l)p_2 \\ & + (1 - H(n_1 + n_2 - l))p_1) \end{aligned}$$

This problem can be solved in two steps:

[Step I]

Solve two nonlinear programming problems

Programming I

$$\begin{aligned} \max_{n_1, n_2} & (\int_0^{n_1} S_1 x dF_1(x) + \int_{n_1}^{+\infty} S_1 n_1 dF_1(x) + \int_0^{n_2} S_2 x dF_2(x) \\ & + \int_{n_2}^{+\infty} S_2 n_2 dF_2(x)) - (n_1 + n_2)p_2 \end{aligned} \tag{1}$$

$$\text{s.t. } \begin{cases} n_1 + n_2 \geq l \\ n_1 \geq 0 \\ n_2 \geq 0 \end{cases}$$

Programming II

$$\begin{aligned} \max_{n_1, n_2} & \int_0^{n_1} S_1 x dF_1(x) + \int_{n_1}^{+\infty} S_1 n_1 dF_1(x) + \int_0^{n_2} S_2 x dF_2(x) \\ & + \int_{n_2}^{+\infty} S_2 n_2 dF_2(x) - (n_1 + n_2) p_1 \end{aligned} \quad (2)$$

$$\text{S.T.} \quad \begin{cases} n_1 + n_2 < l \\ n_1 \geq 0 \\ n_2 \geq 0 \end{cases}$$

[Step II]

Comparing the value of the objective function (1) and (2), the larger one is the optimal solution.

Using this method and based on the following theorems, we can get the retailers' optimal order strategy.

Theorem 1 If $n_1^* + n_2^* \geq l$, the optimal order quantity for retailer 1 is $N_1^* = n_1^*$; and that for retailer 2 is $N_2^* = n_2^*$.

Proof: Because

$$\begin{aligned} & \left(\int_0^{n_1} S_1 x dF_1(x) + \int_{n_1}^{+\infty} S_1 n_1 dF_1(x) + \int_0^{n_2} S_2 x dF_2(x) \right. \\ & \left. + \int_{n_2}^{+\infty} S_2 n_2 dF_2(x) - (n_1 + n_2)(H(n_1 + n_2 - l) p_2) \right. \\ & \left. + (H(n_1 + n_2 - l) p_1) \right) \\ & \leq \left(\int_0^{n_1} S_1 x dF_1(x) + \int_{n_1}^{+\infty} S_1 n_1 dF_1(x) + \int_0^{n_2} S_2 x dF_2(x) \right. \\ & \left. + \int_{n_2}^{+\infty} S_2 n_2 dF_2(x) - (n_1 + n_2) p_2 \right) \end{aligned}$$

we can get that

$$\begin{aligned} \max_{n_1, n_2} & \left(\int_0^{n_1} S_1 x dF_1(x) + \int_{n_1}^{+\infty} S_1 n_1 dF_1(x) + \int_0^{n_2} S_2 x dF_2(x) \right. \\ & \left. + \int_{n_2}^{+\infty} S_2 n_2 dF_2(x) - (n_1 + n_2)(H(n_1 + n_2 - l) p_2) \right. \\ & \left. + (H(n_1 + n_2 - l) p_1) \right) \\ & \leq \max_{n_1, n_2} \left(\int_0^{n_1} S_1 x dF_1(x) + \int_{n_1}^{+\infty} S_1 n_1 dF_1(x) + \int_0^{n_2} S_2 x dF_2(x) \right. \\ & \left. + \int_{n_2}^{+\infty} S_2 n_2 dF_2(x) - (n_1 + n_2) p_2 \right) \\ & = \int_0^{n_1^*} S_1 x dF_1(x) + \int_{n_1^*}^{+\infty} S_1 n_1^* dF_1(x) + \int_0^{n_2^*} S_2 x dF_2(x) \\ & \left. + \int_{n_2^*}^{+\infty} S_2 n_2^* dF_2(x) - (n_1^* + n_2^*) p_2 \right) \end{aligned}$$

since $n_1^* + n_2^* \geq l$, hence

$$\begin{aligned} & \int_0^{n_1^*} S_1 x dF_1(x) + \int_{n_1^*}^{+\infty} S_1 n_1^* dF_1(x) + \int_0^{n_2^*} S_2 x dF_2(x) \\ & + \int_{n_2^*}^{+\infty} S_2 n_2^* dF_2(x) - (n_1^* + n_2^*) p_2 \\ & = \int_0^{n_1^*} S_1 x dF_1(x) + \int_{n_1^*}^{+\infty} S_1 n_1^* dF_1(x) + \int_0^{n_2^*} S_2 x dF_2(x) \\ & + \int_{n_2^*}^{+\infty} S_2 n_2^* dF_2(x) - (n_1^* + n_2^*)(H(n_1^* + n_2^* - l) p_2 \\ & + (H(n_1^* + n_2^* - l) p_1)) \end{aligned}$$

Therefore, it can be deduced that (n_1^*, n_2^*)

$$\begin{aligned} & = \arg \max_{n_1, n_2} \left(\int_0^{n_1} S_1 x dF_1(x) + \int_{n_1}^{+\infty} S_1 n_1 dF_1(x) + \int_0^{n_2} S_2 x dF_2(x) \right. \\ & \left. + \int_{n_2}^{+\infty} S_2 n_2 dF_2(x) - (n_1 + n_2)(H(n_1 + n_2 - l) p_2) \right. \\ & \left. + (H(n_1 + n_2 - l) p_1) \right) \end{aligned}$$

Q.E.D

Lemma 1 The equation group

$$\begin{cases} n_1 + n_2 = l \\ S_1(1 - F_1(n_1)) = S_2(1 - F_2(n_2)) \end{cases}$$

has one and only one solution (z_1, z_2) .

Proof: The existence of the solution is clear. Now, we will prove the uniqueness of the solution. Suppose there are two pairs of solutions $(z_1, z_2), (z_1', z_2')$. Because $z_1 + z_2 = z_1' + z_2' = l$, $z_1 - z_1' = z_2' - z_2$, hence $(F_1(z_1) - F_1(z_1'))(F_2(z_2) - F_2(z_2')) < 0$.

Since

$$\begin{aligned} S_1(1 - F_1(z_1)) &= S_2(1 - F_2(z_2)), \\ S_1(1 - F_1(z_1')) &= S_2(1 - F_2(z_2')), \end{aligned}$$

We can deduce that

$$S_1(F_1(z_1) - F_1(z_1')) = S_2(F_2(z_2) - F_2(z_2'))$$

which contrasts to

$$(F_1(z_1) - F_1(z_1'))(F_2(z_2) - F_2(z_2')) < 0 \quad \text{Q.E.D.}$$

Suppose

$$\begin{aligned} \pi_1 &= \left(\int_0^{z_1} S_1 x dF_1(x) + \int_{z_1}^{+\infty} S_1 z_1 dF_1(x) + \int_0^{z_2} S_2 x dF_2(x) \right. \\ & \left. + \int_{z_2}^{+\infty} S_2 z_2 dF_2(x) - l p_2 \right) \\ \pi_2 &= \left(\int_0^{n_1^*} S_1 x dF_1(x) + \int_{n_1^*}^{+\infty} S_1 n_1^* dF_1(x) + \int_0^{n_2^*} S_2 x dF_2(x) \right. \\ & \left. + \int_{n_2^*}^{+\infty} S_2 n_2^* dF_2(x) - (n_1^* + n_2^*) p_1 \right) \end{aligned}$$

Theorem 2 If $n_1^* + n_2^* < l$ and $\pi_1 > \pi_2$, the optimal order quantity for two retailers is $N_1^* = z_1; N_2^* = z_2$ respectively. Otherwise, the optimal solution is

$$N_1^* = n_1^*; N_2^* = n_2^*$$

Proof: If $n_1^* + n_2^* < l$. Using K-T condition, the optimal solution for Programming I is (z_1, z_2) and the objective value is π_1 . The optimal solution for Programming II is (n_1^*, n_2^*) and the objective value is π_2 . Therefore, we can deduce the result directly.

Q.E.D

Based on the above theorems, the optimal order quantity for retailer i, N_i^* , can be calculated with the following algorithm.

[Step I]

Comparing $n_1^* + n_2^*$ with l , if $n_1^* + n_2^* \geq l$, the optimal retailer 1 is $N_1^* = n_1^*$; and that for retailer 2 is $N_2^* = n_2^*$. Otherwise, we will begin the step II.

[Step II.]

Comparing π_1 with π_2 . If $n_1^* + n_2^* < l$ and $\pi_1 > \pi_2$, the optimal solution is $N_1^* = z_1; N_2^* = z_2$. Otherwise, the optimal solution is $N_1^* = n_1^*; N_2^* = n_2^*$.

4. THE SUPPLIER'S OPTIMAL DISCOUNT STRATEGY

We use a Stackelberg game framework to study the supplier's strategy. When no quantity discount is offered, the supplier's unit selling price is p . Then the supplier's profit is $(p - c)(n_1^* + n_2^*)$. When using GBA, with price curve Q , suppose the retailer i 's optimal order quantity according the former section is $N_i(Q)$, the payoff of the supplier is

$$(H(N_1(Q) + N_2(Q) - l)p_2 + (H(N_1(Q) + N_2(Q) - l) - p_1 - c)(N_1(Q) + N_2(Q)))$$

The supplier will determine the discount price p_2 and discount quantity threshold l .

Assumption 1 Suppose if order N_1 and N_2 will bring the retailer the same expected profit and $N_1 < N_2$, the retailer will choose N_2 because more order quantity implies a higher consumer satisfaction.

The supplier will determine the discount price p_2 and discount quantity threshold l .

Let we define

$$\pi = \int_0^{n_1^*} S_1 x dF_1(x) + \int_0^{n_2^*} S_2 x dF_2(x),$$

$$n_1(l) \text{ and } n_2(l) \text{ s.t. } \begin{cases} n_1(l) + n_2(l) = l \\ S_1(1 - F_1(n_1(l))) = S_2(1 - F_2(n_2(l))) \end{cases},$$

for any given p_2 , suppose the discount threshold is $l(p_2)$ s.t. $l(p_2) > n_1^* + n_2^*$ and

$$\pi = \left(\int_0^{n_1(l)} S_1 x dF_1(x) + \int_{n_1(l)}^{+\infty} S_1 n_1(l) dF_1(x) + \int_0^{n_2(l)} S_2 x dF_2(x) + \int_{n_2(l)}^{+\infty} S_2 n_2(l) dF_2(x) \right) - l(p_2)p_2$$

Lemma 2 For any given p_2 , optimal $l(p_2) \geq l(p_2)$

Proof: If $l < l(p_2)$

Case 1: if $l \leq n_1^* + n_2^*$, the sold quantity is $n_1^* + n_2^*$ and the price is p_2 . The supplier's profit is $(n_1^* p_2 + n_2^* p_2)(p_2 - c) < l^*(p_2)(p_2 - c)$.

Case 2: if $n_1^* + n_2^* < l < l(p_2)$, the sold quantity is l , and the supplier's profit is $l(p_2 - c) < l^*(p_2)(p_2 - c)$.

Hence set $l < l(p_2)$ is not optimal for the supplier.

Q.E.D

If $l > l(p_2)$, the retailer i 's order quantity is n_i^* and the supplier's profit is $(p_1 - c)(n_1^* + n_2^*)$. When $l = l(p_2)$, the supplier's profit is $(p_2 - c)l = (p_2 - c)l(p_2)$. Let $p_2^* = \arg \max_{p_2} (p_2 - c)l(p_2)$. Then the supplier's optimal profit is $(p_2^* - c)l(p_2^*)$.

Theorem 3 If $l = l(p_2)$ and $(p_2^* - c)l(p_2^*) > n_1^* + n_2^*$, the optimal price curve is $(p, p_2^*; l)$; else, trading price will keep p , i.e. the supplier should not use GBA, i.e., do not offer any discount.

Employing the GBA, all the supplier and retailers will get more profit, i.e., it is a Pareto improvement.

5. CONCLUSIONS AND FUTURE RESEARCH

In this paper, we employ the group-buying auction as a special pricing mechanism and study the supplier and retailers' discount strategy and order strategy in the B2B e-Market. After building up the model framework, we optimize the retailers' payoff and deduce the optimal order strategy. Then, from the perspective of the supplier, the optimal discount strategy is also studied. We conclude that with the GBA, all members of a supply chain will improve their profits.

In our model we assume that the retailers are cooperative with each other during the ordering decision. However, if they make the decision completely independently, it is another story. Hence, it is interesting and meaningful to study such a case.

Another important area for the future research is to extend the number of the retailers from 2 to k . It is more practical and helpful for the decision making of the supplier.

ACKNOWLEDGEMENT

This work was supported in part by the National Science

Foundation of China (NSFC) under Grant No. 70329001 and 70321001.

REFERENCES

- [1] Anand, K and R. Aron, "Group-buying on the web: A comparison of price discovery mechanisms", *Management Science*, Vol. 48, No. 10, pp1546-1562, 2003.
- [2] Arcelus, F.J., H.S. Nita and G. Srinivasan, "Retailer's pricing, credit and inventory policies for deteriorating items in response to temporary price/credit incentives", *International Journal of Production Economics*, 81-82: pp153-162, 2003.
- [3] Buscher U., Lindner G., Ensuring feasibility in 'a generalized quantity discount pricing model to increase supplier's profits, The Journal of the Operational Research Society. Oxford: Jun 2004. Vol.55, Iss. 6; pg. 667
- [4] Chen, J., X.L. Chen and X.P. Song, "Bidder's strategy under group-buying auction on the Internet". *IEEE Transaction on Systems, Man and Cybernetics: Part A*, Vol. 32, No. 6, pp680-690, 2002.
- [5] Chen, J., Y.H. Liu and X.P. Song, "Group-Buying Online Auction and Optimal Inventory Policy in Uncertain Market", *Journal of Systems Science and Systems Engineering*, 2004
- [6] Crowther J., Rationale for quantity discounts, *Harvard Bus Rev* (March-April 1964), 121-127.
- [7] Horn, T.V. and N. Gustafsson, *Demand Aggregation through Online Buying Groups*, U.S. Patent 6047266, 2000.
- [8] Kauffman, R.J. and B. Wang, "New buyers' arrival under dynamic pricing market microstructure: The case of group-buying discounts on the Internet", *Proceedings of the 34th Hawaii International Conference on System Sciences*, 2001.
- [9] Kahn, G., "Group buying and the Internet". Webehavior white paper, available at www.webehavior.com/group.htm, 2000.
- [10] Kauffman, R.J., and B. Wang, "Bid together, buy together: On the efficacy of group-buying business models in Internet-based selling". P.B. Lowry, J.O. Cherrington and R.R. Watson(eds), *Handbook of Electronic Commerce in Business and Society*, Boca Raton, FL: CRC Press, 2002.
- [11] Klemperer, P., "Auction theory: A guide to the literature", *Journal of Economic Surveys*, Vol. 13, No. 3, 1999, pp227-286.
- [12] Madigan, S, "A new link to making huge savings", available on the Internet at aca.ninemsn.com.au/stories/1019.asp, 2002
- [13] Monahan J.P., A quantity discount pricing model to increase vendor's profits, *Management Science* 30 (6),.1984, 720-726.
- [14] Myerson, R. B., "Incentive compatibility and the bargaining problem", *Econometrica*, Vol. 47, No. 1, pp61-73, 1979.
- [15] Papachristos S. , Skouri K., An inventory model with deteriorating items, quantity discount, pricing and time-dependent partial backlogging, *International Journal of Production Economics* 83, 2003, 247-256
- [16] Rosenblatt M.J., H.L. Lee, Improving profitability with quantity discounts under "xed demand, *IIE Transactions* 17 (4), 1985 388-395.
- [17] Vickrey, W., "Counterspeculation, auctions, and competitive sealed tenders", *The Journal of Finance*, Vol. 16, No. 1, 1961, pp8-38.
- [18] Wee, H.M., Deteriorating inventory model with quantity discount, pricing and partial backordering, *International Journal of Production Economics* 59 ,1999, 511-518
- [19] Yang P.C., "Pricing strategy for deteriorating items using quantity discount when demand is price sensitive", *European Journal of Operational Research* 157 (2004) 389-397
- [20] Zahir S., Sarker R., Joint economic ordering policies of multiple wholesalers and a single manufacturer with price-dependent demand functions, *Journal of the Operational Research Society* 42 (1991)157-164.

An E-Logistics Model for Effective Collaborative Commerce

Pradeep Kumar, Tapas Mahapatra

ICFAI Business School, ICFAI University, IDPL Complex, Old Delhi Jaipur Road, Dundaheera, Gurgaon (Haryana) 122016, India
{pkgarg, Tapas}@ibsdel.org

ABSTRACT

Collaborative Commerce (C-Commerce) is the name given to commercial relationships carried out over a collaborative framework to integrate enterprises' business processes, share customer relationships and manage knowledge across enterprise boundaries. This is a business strategy that leverages the extensive reach of the Internet. C-Commerce places emphasis on supply chain involvement where members of the supply chain work together to design, build, market and deploy products and services. Collaborative commerce is the next generation of E-commerce framework. E-logistics is the mechanism of automating the logistics processes and providing an integrated, end-to-end fulfillment and supply chain management service to the players of logistics processes. This paper attempts to design an integrated framework for an E-logistics model that it can be effectively employed for C-Commerce.

Keywords: Collaborative Commerce, E-Logistics

1. INTRODUCTION

Collaborative Commerce (C-Commerce) is the name given to commercial relationships carried out over a collaborative framework to integrate enterprises' business processes, share customer relationships and manage knowledge across enterprise boundaries. The ultimate aims of C-Commerce initiatives are to maximize return on intellectual capital investment, business agility and the quality of the customer experience. C-Commerce is a business strategy that leverages the extensive reach of the Internet. C-Commerce places emphasis on supply chain involvement where members of the supply chain work together to design, build, market and deploy products and services. This helps to accelerate the business pace, creates barriers to entry in a particular industry, it enhances customer loyalty, and enables new business models. It is far more crucial than basic B2B e-commerce that is designed to construct a virtual link for a pre-defined community of trading partners to buy or sell goods and services. Even after the fall of the dot.com era, corporate strategies and venture capitalists are embracing C-Commerce as the next generation of e-commerce and an evolution of the traditional supply chain process [1].

There are major driving forces that led to the development of c-commerce. They include expansion of global communications, cost efficiency and economies of scales, customer anticipations of cheaper and better products along with faster delivery and fast access to information, high-speed communication networks. Additionally, some of the driving forces incorporate high-speed communication networks, affordability of business applications, quick deployment, and standardization of technology [2].

The three ingredients for effective C-Commerce are (a)

Managing supply chain, (b) Leveraging technology, and (c) Adapting business models to current market environment. There are six critical factors identified for any organization to successfully evolve from E-commerce to C-Commerce [4]. These factors are:

1. Desire and capacity for fundamental change
2. Internal capabilities for collaboration
3. Tolerance for managed risk
4. Organizational agility
5. Effective implementation
6. Disciplined approach to performance measurement

Today, collaboration is the key factor enabling companies to transcend traditional constraints and perform competitively in a dynamic, increasingly complex business environment [3].

2. NEXT GENERATION E-COMMERCE

Collaborative commerce is the next generation of E-commerce framework. E-commerce is different in the sense that it primarily deals with transaction involving order placement, fulfillment, and payment delivery whereas C-commerce provides dynamic, web-enabled exchange of intellectual capital within enterprise and trading partners to optimize product development, supply chain operations, and manufacturing processes [5].

The Figure-1 above represents how C-Commerce brings people together to make real time decisions across organizational boundaries. It also unites the enterprise in a single secure environment.

2.1 Stages of C-Commerce Evolution

With the birth of the Internet, companies started to do commerce on line with storefront models, from there,

companies started to use the Internet to help make their internal processes better so that they can provide better products and customer service and increase profits by lowering transaction cost using Internet technologies like EDI. The next evolution was the use of the Internet by companies to connect to their suppliers. This allowed for lower inventory cost, better customer service - getting the product there at the right time, and increase in efficiency and productivity from both ends [6].

2.2 Enablement of Multi-Dimensional Collaboration

Web services allow organizations to participate in multi-dimensional collaboration where organizations can collaborate with several partners at a time. This increases inter-enterprise processes and supply-chain velocity. C-commerce is vital in streamlining business process and bringing cost efficiencies and revenue potential for today's organizations, and with the use of Web services integration among business processes is faster and better [7].

2.3 How Web Services Redefine C-Commerce?

So far, we are in a very early stage in the evolution of web business. We have been able to overcome the issue of the standardization of enterprise application programming interfaces around the web, and have managed to undertake tremendous business-to-business exchange transactions over the web, albeit with high integration costs. Web Services promise to revolutionize this process.

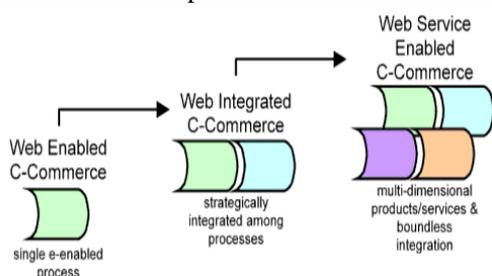


Figure 2

In the figure above, we can see the three stages of c-commerce, from bottom-left to top-right:

First stage of C-Commerce - web enabled c-commerce, a one dimensional, single e-enabled business process that allows certain internal data to be visible to external trading partners, and vice versa. Typically, this meant implementing a web presentation of the data, and allowing partners access to it. This is a limited form of c-commerce, with very limited value, saving only labor cost. Typical applications would be displaying demand for production materials, showing sales forecasts to suppliers, or presenting bills electronically.

Second stage of C-Commerce - B2B exchanges in markets such as steel, auto parts, chemicals or airline equipment. Buyers, sellers and suppliers are integrated

through a web portal. In this stage, each enterprise can reconfigure its supply chain through the marketplace to leverage aggregated buying power and eliminate brokerage fees and middlemen.

The first two stages of C-Commerce have never reached the critical mass required for mass adoption by all industries, because of the cost of integration.

Third stage of C-Commerce - will be built on Web Service as a core integration engine to deliver seamless process integration, seamless customer satisfaction integration and seamless product design integration. It is a plug and play sort of C-Commerce rather than a hard-wired, integration driven effort. A new study by strategy consultants Roland Berger provides an in depth look and guide for the current development of business-to-business (B2B) E-Commerce in the consumer goods industry. Last year both manufacturers and retailers began to invest heavily in B2B initiatives such as the GlobalNetXchange and Transora. For leading industry CEOs and senior executives the key question is how will Internet exchanges affect relationships with their retail customers. Customer relationship management and supply chain processes are, for many consumer goods companies, major sources of differentiation and competitive advantage [8].

3. LATEST GENERATION OF SUPPLY CHAIN MANAGEMENT

The latest generation of supply chain management is Web-Centric. It is characterized by the marriage of the Internet and the supply chain and has resulted in the birth of electronic business (e-business) applications. These Internet enabled, e-business applications have Internet integrated all branches of the supply chain and emerged as the most cost effective means of supply chain operation [9]. E-business applications (e-procurement, e-commerce, and e-collaboration applications) change the supply chain from a linear, rigid chain into a dynamic chain based on an information hub called an ERP (Enterprise Resource Planner), as shown in Figure 3 below.

The Spoke and Wheel Internet-Enabled Supply Chain



Figure 3

When it comes to logistics, the challenge has always been how to deliver products to customers as quickly and safely as possible. Logistics is concerned with the flow of materials in the supply chain, from source through the industrial process to the customer, and then on to re-use, re-cycle, or disposal. By coordinating all resources, logistics have to ensure that service level agreements with customers are honored. Efficient logistics can result in cost savings, which can be passed on to the customer, often resulting in increased business [10].

4. E-LOGISTICS

E-logistics is the mechanism of automating the logistics processes and providing an integrated, end-to-end fulfillment and supply chain management service to the players of logistics processes. Those logistics processes that are automated by e-logistics provide supply chain visibility and can be part of existing e-commerce or workflow systems in an enterprise [13] [14] [16] [18].

The typical e-logistics processes include Request For Quotes (RFQ), Shipping, and Tracking. As shown in the following diagram, e-Logistics interacts with the business process manager in an e-commerce server such as B2B (business to business) or B2C (business to consumer) server, as shown in the Figure 4..

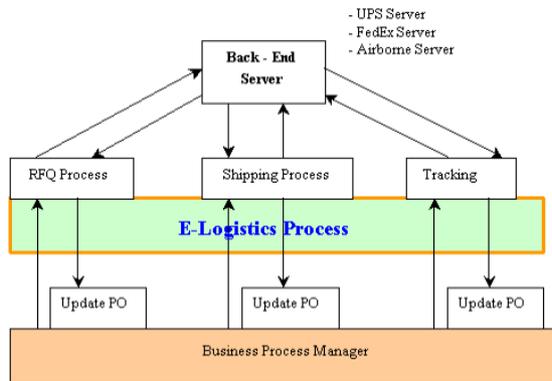


Figure 4

The business process manager invokes the RFQ process to get the basic services such as obtaining the quotes in an e-logistics process. Whenever a response is obtained, the purchase order (PO) will be updated. The shipping process is also invoked by the business process manager and will update the corresponding PO upon

completion. Along with the shipment of goods, a tracking number will be given to the customer and that tracking number will be bound to the PO number in the processing e-commerce system. Customers can track their shipment with the help of that number. The interaction diagram of e-logistics and business process manager shown above represents the high-level view of ELPIF.

5. CONCLUSION

Because the Internet extends around the world, it offers the perfect information sharing channel for logistics. With it, customers can buy and pay for materials, direct their delivery, and trace an order every step of the way. Organizations realize that a strong supporting logistics or electronic logistics (e-logistics) function is an important organizational offering from both the commercial and consumer perspective. The implications of e-logistics models and practices cover the forward and reverse logistics functions of organizations. They also have direct and profound impact from an environmental perspective. This paper tries to come out with a framework to implement C-commerce with the active help of E-logistics.

REFERENCES

1. www.computerworld.com
2. www.businessweek.com
3. www.acm.org
4. collaborationadvisor.com
5. www.intelligententerprise.com
6. www.itworld.com
7. internet.about.com
8. www.pantellos.com
9. logistics.about.com
10. www.e-insite.net
11. www.sagainfotech.com
12. www.helixcommerce.com
13. www.eajournal.com
14. www.fisig.com
15. www.muirmedia.com
16. www.logistics.about.com
17. www.ircbeginner.com
18. www.idea-inc.org

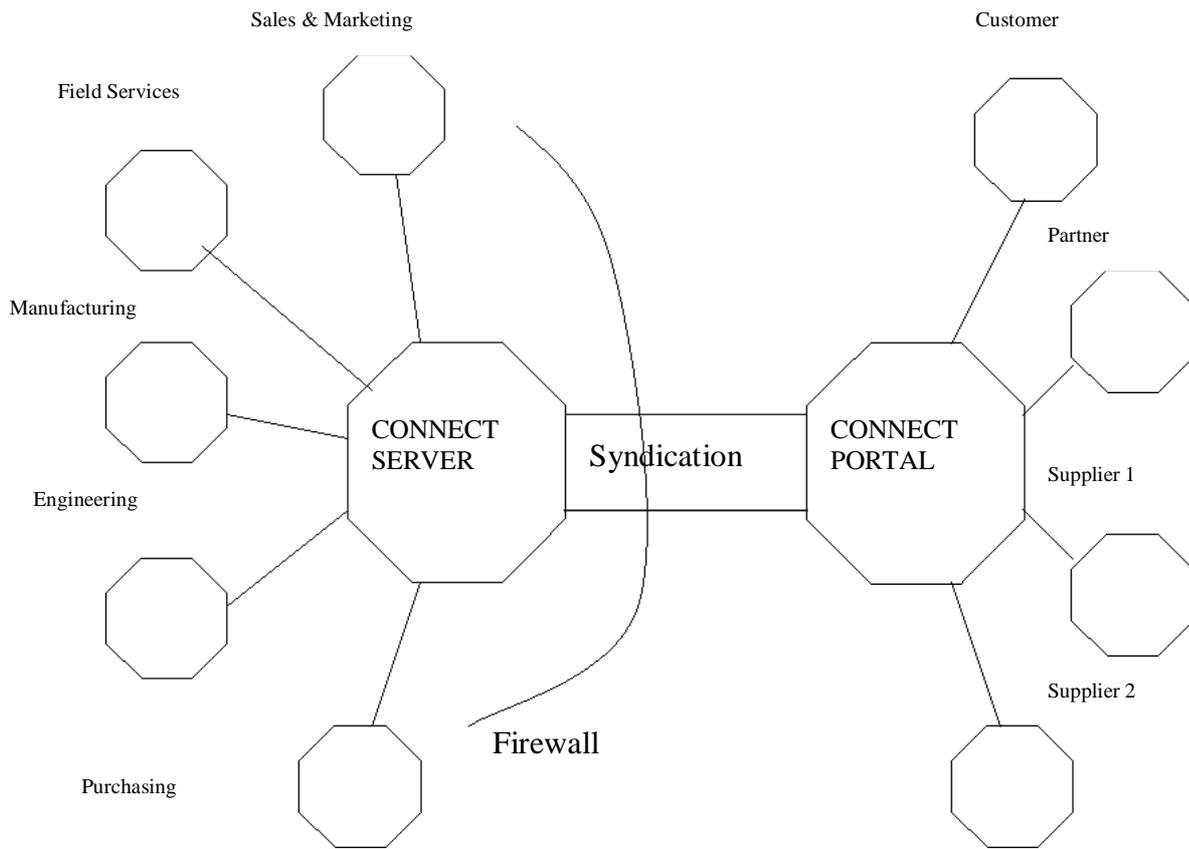


Figure 1

An XCS-Based Intelligent Searching Model for Cross-Organization Identity Management in Web Service

Chichun Lo¹, Weili Lin², Jay Wu³

¹ Professor, Institute of Information Management, National Chiao-Tung University, Hsin-Chu, 300, Taiwan, China, cclo@faculty.nctu.edu.tw

² PhD Student, Institute of Information Management, National Chiao-Tung University, Hsin-Chu, 300, Taiwan, China, weililin.iim92@nctu.edu.tw

³ Associate Engineer, Institute for Information Industry, Taipei, 106, Taiwan, China, jay@iii.org.tw

ABSTRACT

Internet services in distributive organization are normally built on an open network environment. In the environment internet service provisioning cannot be expected executing in a single close organization [1]. Identity management in cross-organization becomes an issue for handling Internet service and distributive business process.

The “identity” in cross-organization web service is defined as global identity rather than private identity from client. Global identity searching table that registers all related service organization is the normal way we used to [2]. Through global identity searching table the target service organization can be looked up directly. For some business program, however, global identity is not necessary registered in specific service organization, e.g. IMSI (International Mobile Subscriber Identification) registration in NP (Number Portability) Service [3], for instance. In NP Service each IMSI can be re-assigned to different mobile service provider if the IMSI apply the re-assignment.

In the example about IMSI in NP service mentioned above, if there is an internet service will be executed according to identity management with IMSI, then it will be a challenge to find the organization for available IMSI in very short response time. To dynamically re-assign a IMSI in different mobile service provider, the traditional global identity searching table will not be practical due to frequently changing the registration of IMSI in different service provider.

To give an intelligent searching model for cross-organization global identity management is a better way than a static global identity searching table management in Web service. In this article the XCS (eXtended Classifier System) classifier system [4] will be proposed as the kernel system. With the characteristics in machine learning and rules management, the XCS-based intelligent searching model can help to predict where the web service can find the global identity in the open and cross-organization environment.

Keywords: Identity Management, Cross-Organization Searching Model, Web Service, XCS Classifier System

1. INTRODUCTION

The cross-organization web service allows a user to seamlessly traverse different sites on different organizations [1]. For any cross-organization service in Internet, global identity is not necessary found from first organization which detects service request from client. It implies that service organization should be “searched” if it will offers required global identity.

In this field a lot of searching scheme under the assumption of fixed network topology, such as Microsoft TrustBridge [6], IBM Tivoli Federated Identity Authentication System and Secure Web Service [7], etc.. Most of major applications for m-commerce or e-commerce, identity management is developed as same way in opening internet environment [1][9].

For any changing in an opening cross-organization web environment, re-configuring of network topology, routing table will be overhead of adjusting global identity distribution. Therefore, the design of an intelligent searching model for getting right service

organization is the major topic in the article.

For achieving the objective, an XCS-Based intelligent searching model for accessing cross-organization global identity is proposed. Not only give the capability to find required global identity in dynamic service organization, but also give capability in searching rules management, learning and GA based evolutionary operations.

2. IDENTITY MANAGEMENT IN CROSS-ORGANIZATION WEB SERVICE

Identity management in cross-organization web service is more complicated than in single organization environment. The major problems are:

1. The table schema for depicting global identity in cross-organization is complicated.
2. The prediction for global identity in unknown organization is not easy to perform.

Before answering the issues mentioned above, some fundamental research should be performed in advanced.

2.1 Cross-Organization Web Service

Web service is the major communication mechanism in e-commerce. Through web service, flexibility to business partners, suppliers and customers is the major inspiration for business operation. Some standards in web service were taken. These standards normally solve the problems of finding, binding and accessing service. The major standards in web service include UDDI (Universal Description, Discovery and Integration), WSDL (Web Service Definition Language) [8], SOAP (Simple Object Access Protocol) [10], SAML [11], etc.

For providing efficient cross-organization web service under the standards above, identity management is necessarily implemented with intelligent way, not only for its feasibility, but also for its flexibility.

2.2 Cross-Organization Identity Management

The challenges to handle identity management in cross-organization web service are:

1. How to get global identity for required application in acceptable response time?
2. How to define the relationship among service organizations?
3. How to predict the address of service organization that the required global identity belong to, especially the required global identity can be "moved" to any service organization?

In web service, normally the scenario of service requested is defined as the following way:

1. A service requester is detected via client ID. Identity management process is started. In the identity management process, related identity document will be analyzed in terms of user's profile, service portfolio, privacy, authorization and transaction. The cache for recent contacting would be adopted for raising access performance.
2. If the global identity is analyzed through identity management and first service organization is hit, a service ticket will be generated and reply to the service requester. Go to 6.
3. Else if the global identity is confirmed that NOT belong to the first service organization, then the searching of right service organization will be run.
4. If right service organization for required global identity is found, then return the global identity back to first service organization, go to 6.

5. If right service organization cannot be found in time, then exceptional handling will be enabled.

6. No further cross-organization searching action will be required. Execute the application service with the delivered global identity.

3. INTELLIGENT XCS MODEL

XCS (eXtended Classifier System) was developed from LCS (Learning Classifier System) [4] and issued by Stewart W. Wilson in 1995. Classifier System [12] is the major research model for evolutionary computation in artificial intelligence. It is an intelligent learning mechanism with rules of expertise, messaging between rules and parallel process. Except the capability derived from LCS, XCS was also derived from related ZCS (Zeroth Level Classifier System) Research since 1994 [13][14] regarding to classifier system simplification.

By the characteristics of rules management, learning capability and evolutionary GA capability, XCS model is very suitable for representing prediction function in changeable environment as what we study in identity management.

3.1 The XCS Interactive Environment

As a classifier system, the interaction between XCS and environment was built on the information which can be depicted as the binary set way as below:

$$\sigma(t) \in \{0,1\}^L \dots\dots\dots (3.1)$$

In equation (3.1), $\sigma(t)$ indicates the detected binary information in timing t , L indicates the length of detected binary information. The action set for handling the detected information $\sigma(t)$ can be expressed as the equation as below:

$$\alpha(t) \in \{a_1, \dots, a_n\} \dots\dots\dots (3.2)$$

In equation (3.2), $\alpha(t)$ indicates the system action for handling $\sigma(t)$ in timing t . Furthermore, for each $\alpha(t)$, the corresponded payoff $\rho(t)$ will be generated. For different condition, the detected $\sigma(t)$ is handled by one-step or multi-steps action. The flag eop will be set if $\sigma(t)$ is processed completely in multi-steps action.

The rp (Reinforcement Program) [18], which was developed by Dorigo and Colombetti in 1998, will be enabled by payoff and eop which is set by XCS for handling $\sigma(t)$ and $\alpha(t)$. Figure 3-1 indicates the interaction among XCS rule set, rp and Environment with related $\rho(t)$, $\sigma(t)$ and $\alpha(t)$ [4]:

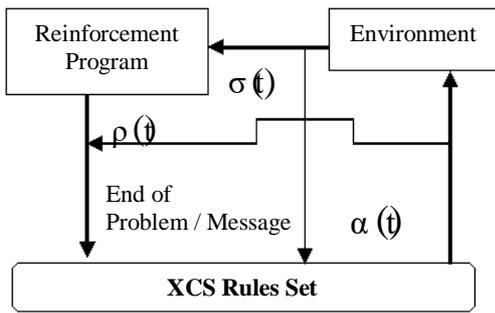


Figure 3-1: The Interaction among XCS Rule Set, Environment and Reinforcement Program with related $\rho(t)$, $\sigma(t)$ and $\alpha(t)$

3.2 XCS Definition and Information Structure

The basic information structure in XCS is derived from normal classifier system which includes Condition, Action and Prediction (State), as table depicted in Figure 3-2:

Classifier									
Condition			Action	Prediction(State)					
1	0	#	1	0	1	0	1	0	1

Figure 3-2: The Basic Information Structure in XCS

However, XCS is more powerful than the traditional classifier system. More system parameters for XCS mechanism including Prediction error ϵ , fitness function f , the counting variable exp of using corresponded XCS classifier in each action, the time stamp ts for last GA operation in action set, the size as of action set and the Numerosity num which is used for counts the numbers of micro classifiers under the appointed macro-classifier.

For any XCS Classifier cl , the attributes can be represented as :

$$x_{cl} \in \{C, A, p, \epsilon, exp, ts, as, num\} \dots \dots \dots (3.3)$$

In addition, XCS is also derived from LCS. According to LCS, the GA operation and learning mechanism in XCS classifier [19] follow MDSE (Markov Decision Support Environment) [20][21][22]. For NMDSE (Non-Markov Decision Support Environment) [23], some extended Internal-State registers were taken for handling learning operation and system evolution.

The XCS system parameters regarding to learning capability includes N (Size of XCS Population), β (Learning Rate), α (Learning Coefficient), ϵ_0 (Precision of ϵ), v (Power Parameter), γ (Factor of Discount), θ_{GA} (GA Threshold), χ (Probability of Cross

Over), μ (Probability of Mutation), θ_{del} (Deletion Threshold), δ (Expected Difference between Fitness and Prediction), θ_{sub} (Threshold of Subsumption), $P_{\#}$ (Probability of Adopting #:Don't Care in condition), p_1 , ϵ_1 , and f_1 for initialize new XCS classifier, p_{explr} (Exploration Probability), θ_{mma} (Threshold to Cover all Conditions [M]), $doGASubsumption$ (Boolean Parameter If Offspring classifier will be tested when subsumption), $doActionSetSubsumption$ (Boolean Parameter If Action Set will be tested when subsumption) that were defined in [4].

3.3 Generalization Hypothesis and System Behavior

Generalization Hypothesis, which was developed by Wilson, is one of the important system features in XCS [4]. The objective of generalization hypothesis is offering of "precise classifier sets" which got complete but non-repeated functional classifiers. The "precise classifier sets" must be as small as possible so that it improves the access performance in traditional classifier system. The Generalization was defined as below[20] :

§ Generalization : is viewed as the problem of finding the smallest set of rules (the classifiers) which completely and accurately represents the task that the agent has solved.

However, generalization is not easy to implement. Because generalization can be viewed as a minimization-covering problem so that normally it was treated as an NP problem [24]. Actually generalization was not a novel model. There were some literatures regarding to artificial intelligence was published for researching how to approach a definite generalization model [23][16]. Of course, XCS is one of the major models in generalization application [5].

With generalization, initialization process and XCS algorithm RUN EXPERIMENT was defined in message handling main loop [4].

According to the algorithm RUN EXPERIMENT, XCS behavior can be depicted as the figure 3-3 shown as below Figure 3-3.

However, for meeting the requirement of generalization, it would make an issue: How can XCS handle new input message $\sigma(t)$ but no right classifier can offer right action?

For the issue mentioned above, 2 operations that can offer "covering" capability and handle the input new message are [4]:

1. Exploration : Actions are selected randomly.

2. Exploitation : always the most predictive action is selected.

For any new classifiers generated via exploration process, it will drive deletion actions for the weakest classifiers in XCS population [P].

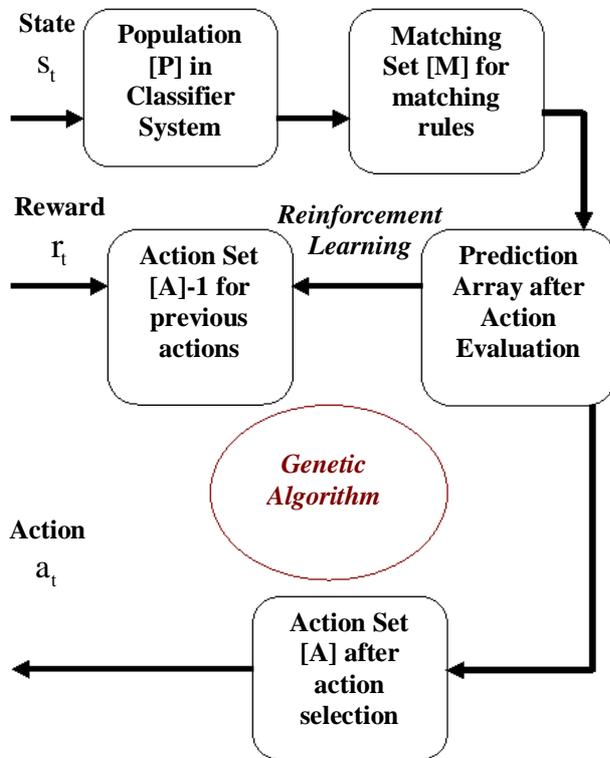


Figure 3-3: The XCS System Behavior

4. XCS-BASED INTELLIGENT SEARCHING MODEL

In cross-organization internet environment, the first service organization which detects the requirement from client is not necessary the one which can offers right global identity for further application process. For the issue from global identity searching in cross-organization web service, XCS got the advantages from its flexible searching rules management, searching rules learning capability and self-evolutionary mechanism via GA operations.

Therefore, both attributes C and A are the key factors to express how to find right global identity for clients service request within input message σ . The information structure of σ is proposed as the data structure shown as figure 4-1:

Input Message σ			
Requested Event ID	Time Stamp	Client ID	Organization A

Figure 4-1: The Info. Structure of Input Message σ

Regarding to input message σ in figure 4-1, the sub-field “Client ID” is the index to find the global

identity in cross-organization web service, sub-field “Organization A” is the target organization for finding global identity (Of course, if the value of “Organization A” is null, then more classifiers in XCS will be invoked to handle the input message σ), the sub-field “Time Stamp” records the execution time which can be taken for verifying if the service scenario is right or not, and the sub-field “Requested Event ID” can be enumerated rely on the requirement of application. For example, in NP telecommunication service, the requested Event ID could be “Login”, “Access Voice Mail”, “Deliver Short Message”, “Turn on service of don’t disturb”, etc., that may be required to execute cross-organization web service in m - commerce environment.

Following the principle of XCS operation, the input message σ will be handled by XCS main loop RUN EXPERIMENT ()[4].

On the other hand, regarding to the condition C in proposed XCS classifier, it can be represented as information in figure 4-2:

Condition C in XCS Classifier x_{cl}			
Condition Operator	Client ID	Organization X	P_Next Condition

Figure 4-2: The Information Structure of Condition C in XCS Classifier

In condition C of XCS classifier x_{cl} , the sub-field “Condition Operator” can be enumerated as the logical operators “NOT”, “OR”, “AND”, “XOR”, etc., that can integrated with more conditions in sub-field “P_Next Condition”. The sub-field “Client ID” is the private identity detected from client. The sub-field “Organization X” is the predicted target organization. The sub-field “P_Next Condition” is the pointer which links to the possible next conditions in the other XCS classifier.

Also, the action A can be represented as information in figure 4-3:

Action A in XCS Classifier x_{cl}			
Action State	Action Type	Organization X	P_Next Action

Figure 4-3: The Information Structure of Condition C in XCS Classifier

In action A of XCS classifier x_{cl} , the sub-field “Action State” can be enumerated as “Action On”, “Action Off”, “Invalid”, “In Action”, etc., the sub-field “Action Type” can be enumerated as “Send”, “Receive”, “Initialize”, etc., that will rely on the given application service requirement. The sub-field “Organization X” is the predicted target organization. The sub-field “P_Next Action” is the pointer which links to more actions if necessary.

After defining the input message σ , information structure for condition C and action A in XCS classifier X_{cl} , the relationship of input message σ and macro view of XCS performance system is looked as the figure 4-4:

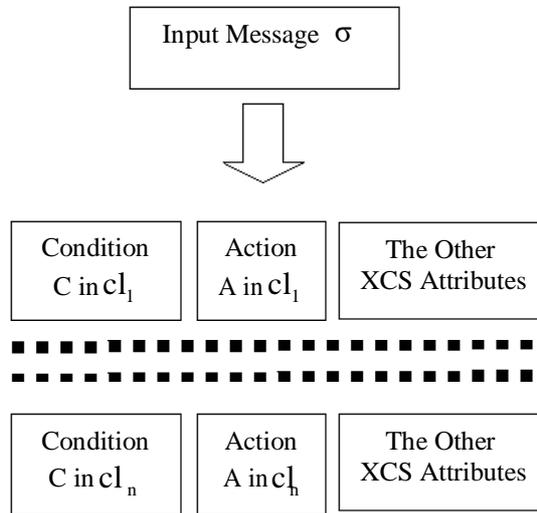


Figure 4-4: The Input Message σ Handling by XCS-Based Intelligent Searching Model

Of course, there are more attributes that can be taken for define global identity such as the personal information as name, sex, birthday, phone number, address, birth place, blood type, etc.. However, the additional personal attributes can be fed back with the returned global identity from the service organization.

5. CONCLUSION

The characteristics of intelligent XCS which offers multi-rules management and GA operation are the foundation of opening environment application. It is why the XCS is proposed for improving identity management in cross-organization web service. Also, the dynamic operation requirement, which cannot be carried out with static global identity searching table, is fulfilled in cross-organization web service.

More cross-organization web services, such as knowledge management, case study in e-business, CRM (Customer Relationship Management), e-learning..., etc., can be implemented with XCS-based intelligent model. However, there are still some issues happen in developing XCS-based intelligent technology in cross-organization web service:

1. For efficiently developing XCS-based intelligent technology, the information defined in input message σ , condition C , action A and the other attributes in XCS classifier, should be represented in semantic way rather than the fixed length format which was used in

traditional classifier system. Such kind of information representation, which is taken for all XCS system parameters, can follow the syntax analysis and semantic analysis executed in compiler or interpreter for program language. To "raise" the information level in XCS-based intelligent model will be beneficial to decode the high layer information representation in cross-organization web service.

2. Through the generalization hypothesis, the classifiers allocated in XCS can be "normally distributed". No repeated classifiers will be allocated in the same XCS. Covering procedure and related GA operation will be executed more efficiently. For improving performance and capability developed in LCS (Learning Classifier System), ZCS (Zeroth level Classifier System) and traditional CS (Classifier System), more than 25 parameters will be handled in the system. Without experiment, potential risk and system shortage cannot be found easily. Therefore, more experiments on dynamic classifier research and simulation should be run as more as possible.

3. For handling higher layer of input message, the neural network can be taken for defining higher level weighted function and neural activation function. With the capability of rules management, learning and GA operation, the XCS can be evolved more efficiently. More application can be developed via this business cooperation.

In the article, XCS was taken for intensifying the identity management in cross-organization web service. The research of innovative artificial intelligent model, such as XCS, is expanded very fast. New theories and methodologies are continually developed. The strength to integrate research with web service and artificial intelligent mechanism, via innovation and considerate analysis, will be the point.

REFERENCES

- [1] Chris Peltz, "Web Services Orchestration and Choreography", IEEE Computer Society, October 2003, Hewlett-Packard Company
- [2] Duncan A. Buell (University of South Carolina), Ravi Sandhu (George Mason University and NSD Security), "Identity Management", IEEE Internet Computing, Nov-Dec, 2003
- [3] Tango Telecom white paper, "Mobile Number Portability – A White Paper from Tango Telecom", Tango Telecom (Ireland), <http://www.tango.ie/pdf/mnp.pdf>, 2000
- [4] Butz, M. and Wilson, S.. In Lanzi, P. L., Stolzmann, W., and S. W. Wilson (Eds.), "An algorithmic description of XCS", Advances in Learning Classifier Systems. Third International Workshop (IWLCS-2000), Lecture Notes in Artificial Intelligence (LNAI-1996). Berlin: Springer-Verlag, 2001
- [5] Gabriel Robert, Pierre Portier, Agnès

- Guillot,"Classifier Systems As 'Animat' Architectures for Action Selection in MMORPG", AnimatLab Laboratoire d'Informatique de Paris, <http://animatlab.lip6.fr/papers/gameon.pdf>, 2003
- [6] Microsoft "Microsoft Windows "TrustBridge" to Enable Organization to Share User Identites Across Business Boundries" ,Press of Microsoft from Radmond, Wash., <http://www.microsoft.com/presspass/press/2002/Jun02/06-06TrustbridgePR.asp>, Jun, 2002.
- [7] IBM, "IBM Tivoli Federated Identity Authentication System and Secure Web Service", IBM Technical White Paper, http://www.rv-nrw.de/Koop/TivoliIdMgr/FederatedIdentity_technicalWP.pdf, Nov, 2002.
- [8] W3C Note, "Web Services Description Language (WSDL) 1.1", <http://www.w3.org/TR/wsdl>, Mar-2001.
- [9] David Clark, "Next-Generation Web Service", IEEE Internet Computing, Mar-Apr 2002.
- [10] W3C Note, "SimpleObject Access Protocol (SOAP) 1.1", <http://www.w3.org/TR/SOAP>, May-2000.
- [11] Von Welch, Rachana Ananthakrishnan, Sam Meder, Laura Pearlman, Frank Siebenlist, "Use of SAML in the Community Authorization Service", <http://www.globus.org/Security/cas/Papers/SAML%20Feedback-aug19.pdf>, Aug-2003.
- [12] J. H. Holland, K. J. Holyoak, R. E. Nisbett, and P. R. Thargard, "Induction: Processes of Inference, Learning, and Discovery", MIT Press, 1986.
- [13] Alwyn Barry, Faculty of Computer Studies and Mathematics, "Specifying Action Persistence within XCS", University of West of England, UK, <http://www.cs.bath.ac.uk/~amb/papers/GECCO2000cs188.pdf>, 2000.
- [14] Jeanan Sha'Aban (Centre for Transportation Studies, University College London), Andy Tomlinson(Intelligent Computer System Centre, University of the West of England), Benjamin Heydecker(Centre for Transportation Studies, University College London), Larry Bull (Intelligent Computer System Centre, University of the West of England), "Adaptive Traffic Control Using Evolutionary Algorithms", http://www.iasi.rm.cnr.it/ewgt/13conference/68_shaaban.pdf, 2003.
- [15] Tim Kovacs, "XCS Classifier System Reliably Evolves Accurate, Complete, and Minimal Representations for Boolean Functions", School of Computer Science, University of Birmingham, UK, <http://www.cs.bris.ac.uk/~kovacs/online97/online97.htm>, 1997.
- [16] Šarūnas Raudys (Vilnius Gediminas Technical University), Masakazu Iwamura (Lithuania Tohoku University), "Multiple Classifier System for Reducing Influences of Atypical Observation", http://www.aso.ecei.tohoku.ac.jp/publication_data/312.pdf, 2003
- [17] Dayle Major, "An Introduction to Learning Classifier Systems", <http://web.umn.edu/~tauritzd/courses/cs447/project/Majors.doc.>, 2002.
- [18] Dorigo. M. & Colombetti. M. "Robot Shaping: An Experiment in Behavior Engineering". Intelligent Robotics and Autonomous Agents. Cambridge, MA, MIT Press, 1998
- [19] C.J. C. H Watkins and P. Dayan, "Q-Learning", Machine Learning, Pg 8:279 ~ 292, <http://citeseer.ist.psu.edu/context/5614/0>, 1992.
- [20] Pier Luca Lanzi, Presentation File – "An Introduction to Learning Classifier Systems", Artificial Intelligence and Robotics Laboratory, Dipartimento di Elettronica e Informazione, Politecnico di Milano, <http://www.elet.polimi.it/upload/lanzi/corsi/icse/2002/Lezione%2010%20-%20Sistemi%20a%20Classificatori.pdf>, 2000
- [21] Abdullah Gani (The University of Sheffield, UK), Omar Zakaria (The Royal Holloway, University of London, UK), Nor Badrul Jumaat (The University of Malaya, Kuala Lumpur, Malaysia), "A Markov Decision Process Model for Traffic Prioritisation Provisioning", Information Science and Information Technology, Pg 905 ~ 913, <http://proceedings.informingscience.org/InSITE2004/117gani.pdf>, 2003
- [22] Ratitch B., Precup D., "Characterizing Markov Decision Process", In 13th European Conf. On Machine Learning (ECML'02), Helsinki, Finland, 2002.
- [23] Pier Luca Lanzi (Artificial Intelligence & Robotics Project, Dipartimento di Elettronica e Informazione, Politecnico), Steward W. Wilson (Prediction Dynamics), "Optimal Classifier System Performance in Non-Markov Environments", <http://world.std.com/~sw/ps/PdMreport99ecj.pdf>, 1999
- [24] Peter Waiganjo Wagacha, "Machine Learning Notes on: I. Classifier Learning and Generalization, II. Data Preparation, III. Validation Methods", Institute of Computer Science, University of Nairobi, http://www.uonbi.ac.ke/acad_depts/ics/course_material/machine_learning/MLNotes.pdf, 2002.

Assessment of Service Quality of Logistics Center in Chain Stores Based on Fuzzy Logic and TOPSIS

Sheng Li

School of Business Administration, Southwestern University of Finance & Economics,
Chengdu 610074, China

ABSTRACT

In this paper, based on analysis of impact factors for service quality of logistics center in chains stores, a set of assessment indexes are proposed. Then, on the basis of this set of assessment indexes, two appraisal models are put forth: one is model based on TOPSIS, the other is fuzzy comprehensive assessment model. The experimental results demonstrate that the two novel methods are effective and feasible.

Keywords: Service quality, logistics center, assessment, ISM, fuzzy logic, TOPSIS

1. INTRODUCTION

If a chain store that wants to success constantly and keeps enhancing core competence, it is necessary to obtain and possess mighty competition fools. Effective logistics management is the key process for the sustainable development of chain stores. For managers usually want to get logistics competition superiority, it is significant to carry out and reinforce logistics management in chain stores. There are commonly seven individual functions in the logistics management of chain store: order, transport, inventory, packaging, warehousing, delivery, processing. And those functions act and impact themselves one another. It is well known that chain management aims at reducing the operation costs, high-efficiency operation and scale economy. The accomplishment of these objectives relied on the effective operation of logistics center. The pursuit for the objectives makes the management of chain stores more complex than that of traditional retail stores. So, it is important for researchers to pay attention to the operation of the logistics center in chain stores.

Service is the core function of the logistics center. In essence, only when free rein is given to the service function, the scale benefits may be reached fully. Thus, it is useful to study the service quality of logistics center. Whether the service quality can be accepted or not is due to customs, and whether they satisfy or not is the ultimate standard to evaluate the service level of the logistics center in a chain store. Satisfactory loyal customs construct the important parts of the core competence and source of benefits of a company. A scientific service quality approach is helpful to get accuracy data, information and the decision foundation and give advices on how to take some countermeasures to enhance service quality. Thus, we think that it is significant to set up a set of assessment indexes of service quality of logistics center and propose a comprehensive appraisal model respectively, which can help us monitor and analyze the operation state and reveal weaknesses of logistics center and give advices on planning and improving of logistics management in

a chain store perfectly.

There are so many impact factors for service level of a logistics center that there are difficulties in selecting the assessment indexes and decide the weights of them. As an effective system analysis tool, ISM (interpretive structural modeling) can help us grasp the chief factors and construct the hierarchy of assessment indexes. After expert survey and consulting, a set of assessment indexes of service quality of logistics center is put forward by using ISM and the corresponding hierarchy structure is set up here. Thus, the assessment problem becomes a multiple criterion decision-making problem in nature. So, what we should do is to establish appraisal model suing multiple attribute decision-making theory. In the work here, we provide two comprehensive appraisal models: one is model based on TOPSIS, the other is fuzzy comprehensive assessment model.

2 .ASSESSMENT INDEXS OF SERVICE QUALITY IN LOGISTICS CENTER

It is dispensable for carry out service quality assessment of logistics center using a scientific appraisal system in order to identify and analyze all the relative impact factors of it and give a corresponding assessment result. Here we propose a set of assessment indexes according to ISM on the basis of comprehensive considering such impact factors as feasibility, economy, hierarchy, time, consistency et al. The steps of ISM are as follows:

- (1) All experts and analysts discuss and select the elements in the system and establish an adjacent matrix.
- (2) Set up reachability matrix.
- (3) Carry out the further study and reasoning on the reachability matrix.
- (4) Conduct the region division and hierarchy division.
- (5) Acquire the reduced reachability matrix.
- (6) Gain the reachability matrix with minimum edges.
- (7) Draw a hierarchy directed graph.
- (8) Make the further analysis and discussion on the studied system.

After the steps abovementioned, we gain a set of assessment indexes of the service performance of a logistics center, which consists of six dimensions: 1) commodity availability; 2) scale economy; 3) information; 4) date of delivery; 5) after-sale service; 6) security. Moreover, each index can be further divided into other sub-indexes. Thus, a hierarchy of assessing service quality of a logistics center is constructed.

The corresponding indexes are listed as follows:

$$U = \{U_1, U_2, U_3, U_4, U_5, U_6\} \quad (1)$$

$$U_1 = \{u_{11}, u_{12}, u_{13}\} \quad (2)$$

$$U_2 = \{u_{21}, u_{22}, u_{23}, u_{24}\} \quad (3)$$

$$U_3 = \{u_{31}, u_{32}, u_{33}, u_{34}\} \quad (4)$$

$$U_4 = \{u_{41}, u_{42}, u_{43}\} \quad (5)$$

$$U_5 = \{u_{51}, u_{52}\} \quad (6)$$

$$U_6 = \{u_{61}, u_{62}\} \quad (7)$$

Where U is the set of assessment indexes for service quality of a logistics center, U_1 is commodity availability, U_2 is scale economy, U_3 is information, U_4 is date of delivery, U_5 is after-sale service and U_6 is security. u_{11} is commodity amount, u_{12} is commodity variety, u_{13} is commodity quality. u_{21} is the total value of commodity delivered, u_{22} is the total delivery cost, u_{23} is cost ratio, u_{24} is the cost of quick response. u_{31} is information convenience, u_{32} is information timeliness, u_{33} is information accuracy, u_{34} is initiative in information releasing. u_{41} is stableness in delivery timely, u_{42} is the ability of quick delivery, u_{43} is the capability in reducing lead time of delivery. u_{51} is returnability and u_{52} is treatment speed of customer complaints. u_{61} is goods safety and u_{62} is personal safety.

3. APPRAISAL MODEL USING TOPSIS

TOPSIS (Technique for order preference by similarity to ideal solution) is a simple and feasible decision method. By introducing ideal solution and negative ideal solution, the method is capable of reflecting the multiple attributions of thing to be appraised and its corresponding appraisal result is consistent with reality.

3.1 Determine the Weights of Assessment Indexes Using AHP

The steps of determining the weights of assessment indexes using AHP are as follows:

Step 1. Construct comparison matrixes

Supposing that there are two sets $U(n), V(m)$. U is assessment item set which contents n elements, V is assessment indexes set which contents m elements. On the basis of analysis of the relationship among the m elements in the set V , we can get the hierarchic relationship of these assessment indexes. As for the indexes in a hierarchy, we can construct the judgment matrixes in each hierarchy according to the criterion they belong to. Here we assume that the judgment matrix is B , $B = (b_{ij})$, where b_{ij} denotes the relative importance degree via pairwise comparison between index i and index j .

Step 2. Computation of weights of assessment indexes and testing consistency

As to each index judgment matrix B , characteristic root method is used to calculate weight vector under a single criterion. Then, carry out consistency testing and stochastic consistency testing of the judgment matrix. If the consistency testing value CI and stochastic consistency testing value CR are all less than 0.1, the judgment matrix is satisfied with the consistency condition, and the weights of indexes are reliable. Otherwise, adjust the judgment matrix and calculate weight vector again.

Step 3. Computation of weights of assessment indexes to the overall objective

On the basis of weight vectors in each hierarchy, by further computation we can get the comprehensive weights of each index with the overall objective.

3.2 Normalization of the Values of Assessment Indexes

According to the evaluation value of n elements in the assessment item set U , we can construct decision matrix A . In spite of the index is profitable type, cost type or range type, it should be transformed into profitable index at first. Thus, we have:

$$A = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} \quad (8)$$

Where x_{ij} is the appraisal value of the i th assessment item with respect to the j th index.

Normalize decision matrix A , we can acquire the normalized decision matrix Z . $Z = (z_{ij})$, $i = 1, 2, \dots, n, j = 1, 2, \dots, m$,

where $z_{ij} = x_{ij} / (\sum x_{ij}^2)^{1/2}$.

3.3 Construct Normalized Matrix

Utilize the weight vector and the normalized decision matrix Z , we have the weighted normalized decision matrix V as follows:

$$V = ZW = \begin{pmatrix} v_{11} & v_{12} & \dots & v_{1m} \\ v_{21} & v_{22} & \dots & v_{2m} \\ \dots & \dots & \dots & \dots \\ v_{n1} & v_{n2} & \dots & v_{nm} \end{pmatrix} \quad (9)$$

Where W is the comprehensive weight matrix of the m indexes by using AHP.

3.4 Decide the Ideal Points for Each Index

According to the weighted normalized decision matrix V , we can determine the positive ideal points and negative ideal points for each index. If index j is the profitable index, we assume:

$$v_j^+ = \max\{v_{ij} \mid i = 1, 2, \dots, n\} \quad j \in J \quad (10)$$

$$v_j^- = \min\{v_{ij} \mid i = 1, 2, \dots, n\} \quad j \in J' \quad (11)$$

Where v_j^+ is the positive ideal solution and v_j^- is the negative ideal solution.

If index j belongs to cost type, we have:

$$v_j^+ = \min\{v_{ij} \mid i = 1, 2, \dots, n\} \quad j \in J \quad (12)$$

$$v_j^- = \max\{v_{ij} \mid i = 1, 2, \dots, n\} \quad j \in J' \quad (13)$$

Where v_j^+, v_j^- have the same definition as in the formulae (10), (11).

3.5 Calculate the Distance to Ideal Points

Calculate the distance to the positive and negative ideal points of each item in the assessment item set U , and the corresponding formulae are as follows:

$$S_i^+ = \sqrt{\sum_{j=1}^m (v_{ij} - v_j^+)^2} \quad i = 1, 2, \dots, n. \quad (14)$$

$$S_i^- = \sqrt{\sum_{j=1}^m (v_{ij} - v_j^-)^2} \quad i = 1, 2, \dots, n. \quad (15)$$

Where S_i^+ is the distance to positive ideal point and S_i^- is the distance to negative ideal point of each assessment item.

3.6 Calculate the Relative Access Degree

Calculate the relative access degree to the ideal points of each item in the assessment item set U , and sequence the elements in U . We have:

$$C_i = \frac{S_i^-}{S_i^+ + S_i^-} \quad i = 1, 2, \dots, n. \quad (16)$$

Where C_i is the relative access degree to the ideal points of each assessment item. According to value of C_i , we can sequence all items to be assessed. The higher C_i is, the better the item is.

4. APPRAISAL MODEL BASED ON FUZZY LOGIC

In the work here, we also formulate the appraisal of the service quality of a logistics center as a fuzzy MADM (multiple attributes decision-making) problem. We use fuzzy comprehensive assessment to achieve the final ranking scores of the service quality of logistics center. Fuzzy comprehensive assessment for a multi-layer system is to carry out comprehensive appraisal from the lowest layer to the uppermost layer step by step until the final assessment results in the uppermost layer is acquired. In the work here, what we should deal with is a three-layer system, the corresponding formulae are as follows:

$$S = W \circ R = W \circ \begin{pmatrix} \omega_1 \circ R_1 \\ \omega_2 \circ R_2 \\ \dots \\ \omega_n \circ R_n \end{pmatrix} \quad (17)$$

Where “o” is matrix product operator, W is the weight vector, and S is the vector of the final assessment result. As to the assessment in the second layer, the computation formula is:

$$R_i = \sum_{j=1}^m \omega_{ij} \circ R_{ij} \quad (18)$$

Where R_i is the assessment score of the factor $U_i (i=1, 2, \dots, n)$ in the hierarchy U , and ω_{ij} is the weight of the factor u_{ij} under the criterion U_i , m is the number of factors u_{ij} belongs to U_i , R_{ij} is the membership vector for the factor u_{ij} , which can be gained by experts judgment and statistical method, and “o” is matrix product operator.

$$S = W \circ R = \sum_{i=1}^n W_i \circ R_i = S (S_1, S_2, \dots, S_m) \quad (19)$$

Where S is the appraisal score of service quality of a logistics center, and W_i is the weight of the factor U_i belongs to U , R_i is the assessment score of the factor U_i in the second layer, “o” is matrix product operator and m is the number of assessment degrees. If $S_k = \max(S_1, S_2, \dots, S_m)$, then the service quality of the logistics center belongs to the degree k .

Namely, according to magnitude of S , we can make conclusions about the service quality of a logistics center. By using maximum membership principle, we can decide which grade a logistics center belongs to and sequence them quantitatively.

In this appraisal model, the weights of assessment indexes can be determined by using AHP as abovementioned.

5. APPLICATION

Based on the analysis process and assessment models abovementioned, we develop a computation procedure for the two models proposed above. Then, we apply this procedure to actual service quality assessment of logistics centers, and the corresponding results are shown in Table 1. The experimental results demonstrate that the assessment conclusion is reliable and is consistent with actual state, and it should be noted that the two models proposed here is feasible and have high precision and their computation results are identical each other.

Table 1. Appraisal score for service quality of logistics center

No.	Logistics center	Model using fuzzy logic		Model using TOPSIS	
		Score	Sequence	Score	Sequence
1	Trust mart	(0.21, 0.52, 0.13, 0.14)	1	0.742	1
2	Hongqi	(0.17, 0.20, 0.47, 0.16)	2	0.657	2
3	Huhui	(0.10, 0.23, 0.44, 0.23)	3	0.601	3
4	Shuangyang	(0.05, 0.19, 0.34, 0.42)	4	0.476	4

6. CONCLUSIONS

It is significant to appraise service quality of a logistics center occasionally, determine assessment indexes and their corresponding weights and get comprehensive evaluation score are indispensable. In this paper, a set of assessment indexes are proposed by using ISM. Then, two appraisal models are put forth: one is model based on TOPSIS, the other is fuzzy comprehensive assessment model. In the two models, the weight of each index is decided by using AHP. The experimental results demonstrate that the two novel methods are effective and feasible.

REFERENCES

[1] Zuo J., *Multi-objective Decision Analysis*,

Hangzhou: Zhejiang University, pp41-89, 1991.

[2] Wang Y. M., "A multi-objective decision and assessment approach: Projecting method", *Study on Statistics*, Vol. 12, No. 4, pp21-25, 1998

[3] Xu S., *Theory of analytic hierarchy process*, Tianjin: Press of Tianjin University, pp43-57, 1988.

[4] Hakimi S. L., "Optimum distribution of switch centers in a communication network and some related graph theoretic problems", *Operation Research*, Vol. 13, pp. 462-475, 1965.

[5] Lai K., "Developing a simulated annealing algorithm for the cutting stock problem", *Computers and Industry*, Vol. 32, No. 1, pp.115-127, 1997.

[6] Robinson E., Laporte, G., "The plant location problem: New models and research prospects", *Operations Research*, Vol. 44, pp. 864-874, 1999.

Asymmetric Demand Information and Channel Profits When Retailers Compete

Bin Xiao¹, Lu Liu², Aling Zhang¹

¹ Institute of Nuclear and New Energy Technology, Tsinghua University, Beijing 100084, China

²School of Economics and Management, Beihang University, Beijing 100083, China
 xiao-b@mail.tsinghua.edu.cn, liulu@buaa.edu.cn, aling@dns.inet.tsinghua.edu.cn

ABSTRACT

We consider a supply chain with a single manufacturer selling a single product through two competing retailers. The manufacturer sets the wholesale price, and the retailers set the retail margin, simultaneously. The demand information among the members is asymmetric. The main contribution of this paper is extended the results of [5] with two competing retailers under asymmetric demand information. We consider six cases with the manufacturer or the retailers whether owning the information or not, and analyze these six cases in turn solve for a Bayesian equilibrium. Some results are obtained finally.

Keywords: supply chain, Channels of distribution, game theory.

1. INTRODUCTION

New market information technique, such as internet developing, B2C electronic commerce application and POS (point of sale) technique etc., making retailer can obtain the demand information of the end customers easily. But manufacturers have long been able to purchase similar information from the information consults company or by conducting marketing research themselves.

We consider a supply chain with a single manufacturer selling its single product through two competing retailers. The main purpose and contribution of this paper is extended the results of [5] with two competing retailers under asymmetric demand information. We examine this phenomenon through a game of asymmetric information played among a manufacturer and two competing retailers.

We denote a as the market scale (i.e., the maximum possible market demand) and assume that there is uncertainty over a . Let a be a random variable taking values in the finite interval $[a_{min}, a_{max}]$, with cumulative distribution function F . The expected value and variance of the variable a are θ, σ^2 respectively. This distribution function is known by all firms. When a firm does not have information about the demand parameter, it only knows F . However, if a firm acquires information, it knows the realization of the random variable, a .

The manufacturer sets the wholesale price w , and the retailers set the retail margin m_1, m_2 , simultaneously. The retail price, p_i , is the sum of the wholesale price and the retail margin and is expressed as

$$p_i = w + m_i \tag{1}$$

Where w is the wholesale price and m_i is the retail margin.

We assume a linear demand function,

$$q_i = a - bp_i + dp_j \tag{2}$$

$(i=1,2, i \neq j, 0 < d < b)$.

Where p_i is the price and q_i is the quantity demanded of the retailer i .

We also assume constant average cost of production, and cost of distribution, which are normalized to zero without loss of generality. All decision-makers are profit-maximizers and assumed to be risk-neutral.

The profit functions of the manufacturer (Π) and the retailer i (π_i) are derived by multiplying respective margins by quantity sold. They are, respectively

$$\begin{aligned} \Pi &= \sum_{i=1}^2 w(a - bp_i + dp_j) \\ &= w(2a - b - d)(2w + m_1 + m_2) \end{aligned} \tag{3}$$

$$\begin{aligned} \pi_i &= m_i(a - bp_i + dp_j) \\ & \quad i=1,2, i \neq j \end{aligned} \tag{4}$$

Note that the two retailers have symmetric profit functions.

2. THE INTEGRATED BENCHMARK AND SYSTEM PROFIT

To establish a performance benchmark, we analyze the problem of an integrated firm. Now, if this supply chain is centrally owned and controlled, a vertically integrated manufacturer obtains the totality of channel profit: therefore, it will seek to maximize channel profit.

$$\begin{aligned} \Pi_{\bar{}} &= p_1q_1 + p_2q_2 \\ &= p_1(a - bp_1 + dp_2) + p_2(a - bp_2 + dp_1) \end{aligned} \tag{5}$$

The problem is concave, and table 1 shows the optimal profit.

3. EQUILIBRIUM ANALYSIS UNDER ASYMMETRIC DEMAND INFORMATION

We consider six cases: (1) case 1: None of the members knows the market scale a ; (2) case 2: All of the members know the market scale a ; (3) case 3: Only the Manufacturer knows the market scale a ; (4) case 4: Only the Retailers know the market scale a ; (5) case 5: Both the Manufacturer and the Retailer 1 know the market scale a , but Retailer 2 do not know; (6) case 6: Only the Retailer 1 knows the market scale a . We analyze these six cases in turn solve for an equilibrium.

3.1 Case 1: None of the members knows the market scale a

In this case, although none of the channel members knows the exact value of the demand parameter, a , each maximizes expected profit taking into account the distribution of a . In particular, the expected profits of the manufacturer is given by

$$E[\Pi] = E[w(a - b(-d - m_1 - m_2))] \quad (6)$$

Differentiating equation (6) with respect to w and taking expectation over a gives the first-order condition below.

$$2\theta - b(-d - m_1 - m_2) = 0 \quad (7)$$

The expected profits of the retailer i is given by

$$E[\pi_i] = E[m_i(a - b(w + m_i) + d(w + m_j))] \quad i=1,2, i \neq j \quad (8)$$

Differentiating equation (8) with respect to m_i and taking expectation over a gives the first-order condition below.

$$\theta - b(w + 2m_i) + (d - w) = 0 \quad i=1,2, i \neq j \quad (9)$$

Since we have a symmetric case $m_1 = m_2$ in equilibrium, (8) and (9) gives

$$w = \frac{\theta b}{(3b - d)(b - d)} \quad (10)$$

$$m_1 = m_2 = \frac{\theta}{3b - d} \quad (11)$$

Substituting (10) and (11) into (6) and (8), we obtain the expected profits of the three members, see table 1.

3.2 Case 2: All of the members know the market scale a .

In this case, all of the channel members know the demand parameter. The first-order condition of the manufacturer is given by

$$2a - b(-d - m_1 - m_2) = 0 \quad (12)$$

The first-order condition of the retailer i is given by

$$a - b(w + 2m_i) + (d - w) = 0 \quad (13)$$

By symmetry $m_1 = m_2$. Solving (12) and (13) gives

$$w = \frac{ab}{(3b - d)(b - d)} \quad (14)$$

$$m_1 = m_2 = \frac{a}{3b - d} \quad (15)$$

Substituting (14), and (15) into (6), and (8) gives and taking expectation over a gives the expected profits (table 1)

3.3 Case 3: Only the Manufacturer knows the market scale a

This more complex case is a game of asymmetric information in which the manufacturer knows the demand scale, a , but the retailers only knows the distribution of a . Accordingly, a strategy for the manufacturer consists of a function $w(a)$, where wholesale price is contingent on the realized value of a , since the retailer does not observe a , the strategy for the retailer i consists of a fixed value of m_i . We are solving for a Bayesian equilibrium[5].

Since the manufacturer chooses a $w(a)$ for each value of a , it maximizes (6). The first-order condition is given by

$$w(a) = \frac{a}{2(b - d)} - \frac{m_1 + m_2}{4} \quad (16)$$

Since the retailers know that the manufacturer's strategy is a function $w(a)$, the retailer i chooses m_i to maximize expected profit

$$E[\pi_i] = E[m_i(a - b(w(a) + m_i) + d(w(a) + m_j))] \quad (17)$$

Substituting (16) into (17) and take the first-order condition gives

$$w = \frac{a}{2(b - d)} - \frac{\theta}{2(3b - d)} \quad (18)$$

It is not difficult to obtain the profits of the three members(table 1).

Similarly, We analyze the other three cases in turn solve for a Bayesian equilibrium and obtain table 1.

Table 1 Individual and System Expected Profits

	E[Π]	E[π ₁]	E[π ₂]	System profit
Integrated channel	-	-	-	$\frac{\theta^2}{2(b-d)} + \frac{\sigma^2}{2(b-d)}$
Case 1	S	T	T	S-2T
Case 2	$S = \frac{2b^2\sigma^2}{(3b-d)^2(b-d)}$	$T = \frac{b\sigma^2}{(3b-d)^2}$	$T = \frac{b\sigma^2}{(3b-d)^2}$	$S-2T = \frac{2b(2b-d)\sigma^2}{(3b-d)^2(b-d)}$
Case 3	$S = \frac{\sigma^2}{2(b-d)}$	T	T	$S-2T = \frac{\sigma^2}{2(b-d)}$
Case 4	S	$T = \frac{b\sigma^2}{(2b-d)^2}$	$T = \frac{b\sigma^2}{(2b-d)^2}$	$S-2T = \frac{2b\sigma^2}{(2b-d)^2}$
Case 5	$S = \frac{\sigma^2}{2(b-d)}$	T	T	$S-2T = \frac{\sigma^2}{2(b-d)}$
Case 6	S	$T = \frac{\sigma^2}{4b}$	T	$S-2T = \frac{\sigma^2}{4b}$

Where $S = \frac{2b^2\theta^2}{(3b-d)^2(b-d)}$, $T = \frac{b\theta^2}{(3b-d)^2}$

4. RESULTS

We denote the expected profit of the manufacture under case *i* as $E[\Pi]_{ci}^*$, and the expected profit of retailer *j* under case *i* as $E[\pi_j]_{ci}^*$. We have following results

Result 1 (1) $E[\Pi]_{c1}^* = E[\Pi]_{c4}^* = E[\Pi]_{c6}^* < E[\Pi]_{c2}^* < E[\Pi]_{c5}^* < E[\Pi]_{c3}^*$

(2) $E[\pi_1]_{c1}^* = E[\pi_1]_{c3}^* < E[\pi_1]_{c2}^* < E[\pi_1]_{c5}^* < E[\pi_1]_{c6}^* < E[\pi_1]_{c4}^*$

(3) $E[\pi_2]_{c1}^* = E[\pi_2]_{c3}^* = E[\pi_2]_{c5}^* = E[\pi_2]_{c6}^* < E[\pi_2]_{c2}^* < E[\pi_2]_{c4}^*$

If retailer *i* has the full information about *a* and can report the information to retailer *j* without any cost, then retailer *i* desires to tell retailer *j* about *a* to increase own profit when the manufacture has no information concerning *a*. When the manufacture has the full information about *a*, retailer *i* has no incentive to do so.

Result 2 (1) When the manufacturer is unaware of *a*, its expected profit $E[\Pi]^* = \frac{2\theta^2 b^2}{(3b-d)^2(b-d)}$.

(2) When Retailer *i* is unaware of *a*, its expected profit

$$E[\pi_i]^* = \frac{\theta^2 b}{(3b-d)^2}$$

Result 3 Total channel profits are highest when the supply chain is vertically integrated; the next in order for only the manufacturer knows the market information. While all members do not know the information, the system profit is minimum.

Although the manufacturer is not a leader in the game, Result 2 expresses the manufacturer to be placed in the predominant position in the supply chain.

Result 4 (1) When the supply chain is vertically integrated, the expected distribution quantity is $q_1+q_2 = \theta$. Otherwise, the expected distribution quantity is

$q_1=q_2 = \frac{\theta b}{3b-d}$ (2) When the supply chain is vertically

integrated, the expected price is $p_1^* = p_2^* = \frac{\theta}{2(b-d)}$.

Otherwise, the expected price is

$$p_1 = p_2 = \frac{\theta(2b-d)}{(3b-d)(b-d)}$$

Since $0 < d < b$ and $\frac{2\theta b}{3b-d} < \theta$, when the supply chain is vertically integrated, the expected distribution quantity is maximum.

REFERENCES

[1] Corbett C, Tang C. Designing supply contracts: Contract type and information asymmetry. *Taylor S, Magazine M, Ganeshan R. Quantitative models for supply chain management*. Boston: Kluwer Academic Publishers, 269-297, 1999.

[2] Tirole J. *The theory of industrial organization*, Cambridge, MA: The MIT Press, 1988.

[3] Choi S C. Price Competition in a Channel Structure with a Common Retailer, *Marketing Science*, Vol 10, pp271-296, 1991.

[4] Choi S C. Price competition in a duopoly common retailer channel. *Journal of Retailing*, Vol 72 No.2, pp117-134, 1996.

[5] Chu Wujin, Messinger P R. Information and Channel Profits. *Journal of Retailing*, Vol 73, NO.4, pp487-499, 1997.

[6] Ingene C, Parry M. Channel Coordination When Retailers Compete. *Marketing Science*, Vol 14, pp360-377, 1995.

[7] Ingene C, Parry M. Manufacturer-Optimal Wholesale Pricing When Retailers Compete. *Marketing Letters*, Vol.9, No 1, pp65-77, 1998.

Bidding Strategies Analysis for Procurement Combinatorial Auctions

J. Chen, He Huang

¹ School of Economics and Management, Tsinghua University, Beijing 100084, China
chenj@em.tsinghua.edu.cn, Huangh02@mails.tsinghua.edu.cn

ABSTRACT

Based on the work of Krishna and Rosenthal (1996) about combinatorial auctions bidding equilibrium analysis and Che(1993)'s research about one-unit multi-attributes auctions, we construct a multi-attributes procurement simultaneous auction for 2 objects, through a first-score(price), sealed-bid format. There are two kinds of bidders: *simple bidders* and a *diverse bidder* considered in this model. Each *simple bidder* is interested in only one of the objects, while the *diverse bidder* is interested in both. With some further assumptions, we finally obtain the equilibrium bidding strategies for the both two kinds of bidders.

Keywords: Bidding Strategies, Procurement, Combinatorial Auctions

1. INTRODUCTION

Many auctions involve the sale (or purchase) of a variety of distinct assets. Examples are auctions for airport time slots, delivery routes and FCC radio spectrum rights. Because of complementarities (or substitution) between the auctioned assets, bidders have preferences not just for particular items but for sets or bundles of items. For this reason, economic efficiency is enhanced if bidders are allowed to bid on bundles or combinations of different assets. Possibly because of FCC combinational auction, auctions where bidders submit bids on combinations have recently received much attention [Rothkopf et. al.(1998), Sandholm (1999)]. Actually, such auctions were proposed as early as 1976, by Jackson, for radio spectrum rights. Increases in computing power have made them more attractive to implement. The value of assets for a particular bidder, may depend strongly on which other assets she wins. Because of the possibility of such synergy, the auction designers require attentions to allowing bids not just for individual asset, but also bids for the combinations of assets, which is called combinational auctions (CAs). When an auctioneer procure items by auctions, she may allow the potential suppliers to bid for various products bundles, which is called Procurement Combinatorial Auctions. It will make variety multi-units procurement much fairer, more feasible and more efficient.

One of the most challenging problems in Procurement CAs is Winners Determination Problem (WDP). The general allocation problem in combinational auctions can be commonly formulated as an integer programming problem which is computationally intractable. Literature about CAs mechanism design is quite rich [e.g., Krishna and Rosenthal, 2000; Parkes, 2001,]. Cantillon and Pesendorfer (2003) consider the problem of identification and estimation in the first price multi-unit auction. The topics on one-unit auctions' bidding strategies and winning probability are abundantly addressed. However, discussions about CAs bidding strategies are inadequate. Krishna and Rosenthal (1996) considered situations where multiple

objects are auctioned simultaneously by means of a second-price sealed-bid auction, and derived the bidding equilibrium strategies for the synergy bidder. Rosenthal and Wang (1996) studied a simultaneous-auction model with synergies, common values, and constructed strategies in which bidders of different types randomize over different bid intervals, also provided necessary and sufficient conditions for such equilibria strategies. Che (1993) studies design competition in procurement by developing a model of two-dimensional (price and quality) auctions, and corresponding bids are evaluated by a scoring rule designed by auctioneer. In this paper, we will discuss bidding strategies in a specific multi-attributes procurement CA setting.

2. MODEL AND ITS ASSUMPTIONS

In this model, there are 2 objects to be procured by an auction simultaneously through a first-price, sealed-bid format. And two kinds of bidders, i.e. *simple bidder* and *diverse bidder* are considered. Each *simple bidder* is interested in only one of the objects, while the *diverse bidder* is interested in two objects. For each of the 2 objects, there are n interested *simple bidders*. Each *simple bidder* has a privately known supply cost function for the object.

Other assumptions and their mathematical expressions are below:

(1) Bidding vector is $\{p, q\} \in R_+^2$, p and q denote bidding price and supply quality separately. Especially, quality, in our model, is a unified number to measure all non-price factors of the objects. The scoring function of the bidding vectors has quasi-linear form

$$S(q, p) = s(q) - p \quad (1)$$

where $s(\cdot)$ is strictly increasing with respect to q .

(2) Number of *simple bidders* for each object is n , only one *diverse bidder* in the basic model.

(3) The cost functions of *simple bidders* are

$$c(q, \theta_i) = \begin{cases} c_1(q, \theta_i), & \text{if bid item 1} \\ c_2(q, \theta_i), & \text{if bid item 2} \end{cases}$$

where θ_i denotes type of *simple bidder* i , which is private information of bidder i , and θ_i is drawn from interval $[\underline{\theta}, \bar{\theta}]$ with identical distribution $F(\cdot)$ which are common knowledge. $c_1(\cdot, \cdot)$ and $c_2(\cdot, \cdot)$ are increasing with respect to q and θ_i . Type parameter of the *diverse bidder* is denoted as θ which is drawn from interval $[\underline{\theta}, \bar{\theta}]$ with the identical distribution $F(\cdot)$, θ is also private information, the cost functions of the *diverse bidder* are $c_1(q, \theta)$ for object 1 and $c_2(q, \theta)$ for object 2.

(4) Profit of *simple bidder* i , with bid $\{p, q\} \in R_+^2$, is denoted as

$$\pi_i(q, p) = p - c_r(q, \theta_i), r=1,2 \quad (2)$$

Profit of the *diverse bidder*, with bid $\{p, q\} \in R_+^2$, is

$$\begin{cases} p_1 - c_1(q, \theta), & \text{get item 1} \\ p_2 - c_2(q, \theta), & \text{get item 2} \\ p_1 - c_1(q, \theta) + p_2 - c_2(q, \theta) + \alpha, & \text{get both} \end{cases} \quad (3)$$

where α is a non-negative constant which is *diverse bidder's* private information.

3. BIDDING STRATEGY ANALYSIS

Let $S_0(\theta) \equiv \max[s(q) - c(q, \theta)]$ for all $\theta \in \theta_i$, by using envelope theorem, we know $S_0(\cdot)$ a strictly decreasing function, therefore its inverse function exists. Let $v \equiv S_0(\theta)$, v can be regarded as scoring capability of supplier. Let $H(v) \equiv 1 - F(S_0^{-1}(v))$, then distribution $F(\cdot)$ of θ is transformed into distribution $H(\cdot)$ of v , which also imply that the one-to-one mapping from v to θ .

Let $b \equiv S(q_s(\theta), p)$, where $q_s(\theta) = \arg \max[s(q) - c(q, \theta)]$, and b is the score of the bidding by supplier with type parameter θ . Let $B(\cdot)$ denote the scoring function with the independent variable v , and $B(\cdot)$ is strictly increasing with v . Note that $p - c(q(\theta), \theta) = v - b$, then we have the expected revenue of any *simple bidder* with type θ_i and quality bidding $q_s(\theta)$:

$$\begin{aligned} \pi(q(\theta_i), p | \theta_i) &= [p - c(q(\theta_i), \theta_i)] \\ &\times \text{prob}\{\text{win} | S(q(\theta_i), p)\} \\ &= [v - b] \{H(B^{-1}(b))\}^n \end{aligned} \quad (4)$$

where $\text{prob}\{\text{win} | S(q(\theta_i), p)\}$ represents the winning probability of bidder whose bidding is $\{q(\theta_i), p\}$.

Therefore, Any *simple bidder* with type θ_i has the following bidding equilibria strategy [Riley and Samuelson, 1981; Che, 1993,]:

$$\begin{cases} q(\theta_i) = \arg \max[s(q) - c_r(q, \theta_i)], & r=1 \text{ or } 2 \\ p(\theta_i) = c(q, \theta_i) + \int_{\theta_i}^{\bar{\theta}} c(q(t), t) \left[\frac{1 - F(t)}{1 - F(\theta_i)} \right]^n dt \end{cases} \quad (5)$$

It is convenient to define

$$L(b) = \{H(B^{-1}(b))\}^n \quad (6)$$

Intuitively, $L(b)$ is the probability of *diverse bidder* to win item 1 (or item 2) with score b , that is, the probability of all the *simple bidders* who are interested in item 1 (or item 2) get score less than b .

We've already known the *simple bidder* will use the bidding strategy in (5). Therefore, the expected revenue of *diverse bidder* with type θ is

$$\begin{aligned} \pi(S_1(q_1(\theta), p_1), S_2(q_2(\theta), p_2) | \theta) \\ = L(S_1) L(S_2) [(p_1 - c_1) + (p_2 - c_2) + \alpha] \\ + L(S_1)(1 - L(S_2))(p_1 - c_1) \\ + L(S_2)(1 - L(S_1))(p_2 - c_2) \end{aligned} \quad (7)$$

In the above equation, first term is the expected revenue of winning both items, while the last two terms separately corresponds to expected revenue of winning only item 1 or only item 2.

Lemma 1. The *diverse bidder* with type θ follows the bidding strategy whose quality-bid $(q_{1s}(\theta), q_{2s}(\theta))$ respectively is

$$\begin{aligned} q_{1s}(\theta) &= \arg \max[s(q) - c_1(q, \theta)] \\ q_{2s}(\theta) &= \arg \max[s(q) - c_2(q, \theta)] \end{aligned} \quad (8)$$

Proof. Assume there exists another equilibrium strategy $\{(q_1, p_1), (q_2, p_2)\}$, and $q_1 \neq q_{1s}, q_2 \neq q_{2s}$,

Then construct a bid $\{(q_1', p_1'), (q_2', p_2')\}$, where

$$\begin{aligned} q_1' &= q_{1s}, \\ q_2' &= q_{2s}, \\ p_1' &= p_1 + s(q_{1s}) - s(q_1), \\ p_2' &= p_2 + s(q_{2s}) - s(q_2) \end{aligned} \quad (9)$$

Note

$$\begin{aligned} S(q_1', p_1') &= S(q_1, p_1) = S_1, \\ S(q_2', p_2') &= S(q_2, p_2) = S_2 \end{aligned} \tag{10}$$

Then, with bid $\{(q_1', p_1'), (q_2', p_2')\}$, the expected revenue of diverse bidder is below,

$$\begin{aligned} &\pi((q_1'(\theta), p_1'), (q_2'(\theta), p_2') | \theta) \\ &= L(S_1) L(S_2) \{ (p_1 - c_1(q_1) + [s(q_{1s}) - c_1(q_{1s}) \\ &\quad - (s(q_1) - c_1(q_1))] + (p_2 - c_2(q_2) \\ &\quad + [s(q_{2s}) - c_2(q_{2s}) - (s(q_2) - c_2(q_2))] + \alpha \} \\ &\quad + L(S_1)(1 - L(S_2)) \{ (p_1 - c_1(q_1) \\ &\quad + [s(q_{1s}) - c_1(q_{1s}) - (s(q_1) - c_1(q_1))] \} \\ &\quad + L(S_2)(1 - L(S_1)) \{ (p_2 - c_2(q_2) \\ &\quad + [s(q_{2s}) - c_2(q_{2s}) - (s(q_2) - c_2(q_2))] \} \end{aligned} \tag{11}$$

Note

$$q_{1s}(\theta) = \operatorname{argmax}_q s(q) - c_1(q, \theta)$$

$$\text{and } q_{2s}(\theta) = \operatorname{argmax}_q s(q) - c_2(q, \theta)$$

therefore,

$$\begin{aligned} &\pi((q_1'(\theta), p_1'), (q_2'(\theta), p_2') | \theta) \\ &> L(S_1) L(S_2) [(p_1 - c_1(q_1)) \\ &\quad + (p_2 - c_2(q_2)) + \alpha] + L(S_1)(1 - L(S_2))(p_1 - c_1(q_1)) \\ &\quad + L(S_2)(1 - L(S_1))(p_2 - c_2(q_2)) \end{aligned} \tag{12}$$

Thus, we have

$$\begin{aligned} &\pi((q_1'(\theta), p_1'), (q_2'(\theta), p_2') | \theta) \\ &> \pi((q_1(\theta), p_1), (q_2(\theta), p_2) | \theta) \end{aligned} \tag{13}$$

This completes the proof. □

From LEMMA 1 and (7), we know the optimization problem of diverse bidder with type θ can be expressed as follows:

$$\begin{aligned} &\max_{p_1 p_2} \pi((q_{1s}(\theta), p_1), (q_{2s}(\theta), p_2) | \theta) \\ &= \max_{p_1 p_2} \{ L(s(q_{1s}(\theta)) - p_1) L(s(q_{2s}(\theta)) - p_2) \\ &\quad [(p_1 - c_1(q_{1s}(\theta))) + (p_2 - c_2(q_{2s}(\theta))) + \alpha] \\ &\quad + L(s(q_{1s}(\theta)) - p_1)(p_1 - c_1) \\ &\quad \times (1 - L(s(q_{2s}(\theta)) - p_2)) \\ &\quad + L(s(q_{2s}(\theta)) - p_2)(p_2 - c_2) \} \\ &\quad \times (1 - L(s(q_{1s}(\theta)) - p_1)) \end{aligned} \tag{14}$$

Generally, the pairwise point (p_1, p_2) which satisfies

the first order condition and the second order condition of (14), is the solution to the maximal expected revenue. Now restrict attention to equal-bid for two identical objects, that is, item 1 and item 2 are the same. Naturally, we can let

$$\begin{aligned} q_{1s}(\theta) &= q_{2s}(\theta) = q_s(\theta) \\ \text{and } c_1(q_s(\theta), \theta) &= c_2(q_s(\theta), \theta) \end{aligned}$$

By LEMMA 1, the quality-bid, $q_s(\theta)$, in equilibrium strategy is uniquely determined by θ , then the corresponding cost $c(q_s(\theta), \theta)$ is also determined by θ uniquely. Now, define optimal cost of bidder with type θ as

$$c^*(\theta) = c_1(q_s(\theta), \theta) = c_2(q_s(\theta), \theta)$$

Thus rewrite (14) as

$$\begin{aligned} &\max_p \pi((q_s(\theta), p), (q_s(\theta), p) | \theta) \\ &= \max_p \{ (L^2(s(q_s(\theta)) - p) \times (2(p - c^*(\theta)) + \alpha)) \\ &\quad + 2L(s(q_s(\theta)) - p)(p - c^*(\theta)) \} \\ &\quad \times (1 - L(s(q_s(\theta)) - p)) \end{aligned} \tag{15}$$

Without loss of generality, let $s(q_s(\theta))$ equal 1, and assume price p is drawn from $[0, 1]$. Define

$$\begin{aligned} &G(\cdot) \equiv H(B^{-1}(\cdot)) \text{ and rewrite (15) as} \\ &\max_p \pi((q_s(\theta), p), (q_s(\theta), p) | \theta) \\ &= \max_p \{ (G^{2n}(1 - p) \times (2(p - c^*(\theta)) + \alpha)) \\ &\quad + 2G^n(1 - p)(1 - G^n(1 - p))(p - c^*(\theta)) \} \end{aligned} \tag{16}$$

By first derivative of (16) with respect to p , we have

$$\frac{d\pi}{dp} = -\alpha G^n(1 - p) + p \frac{1}{n} G(1 - p)^{\theta} \theta c^*(\theta) = 0 \tag{17}$$

Now considering the relationship between the optimal cost of diverse bidder with type θ , i.e., $c^*(\theta)$, and her price bid p , from (17) we have

$$c^*(\theta) = \alpha G^n(1 - p) - p \frac{1}{n} G(1 - p) \tag{18}$$

Actually, (18) is the correspondence from optimal cost $c^*(\theta)$ to optimal price bid p , and we can abstractly

express (18) as $c^*(\theta) = c_p(\theta)$. The only special feature used in the argument is that function $c_p(\cdot)$ convex.

Generally, convexity of $c_p(\cdot)$ is not easy to characterize in terms of the primitive assumptions of our model. However, if $G(\cdot)$ is uniform distribution, i.e., the distribution of production capability v is uniform, then the function $c_p(\cdot)$ is indeed convex. If $c_p(\cdot)$ is not convex, the situation becomes more complicated and not worth pursuing here. Therefore, there exist at most 2 roots of equation (18).

Corresponding to different optimal cost $c^*(\theta)$, denote $p(c^*(\theta))$ as the solutions of (17). Define $p^*(c^*(\theta))$ as the unique root or greater one of the two roots, i.e.,

$$p^*(c^*(\theta)) = \begin{cases} p(c^*(\theta)), & \text{for unique solution} \\ \max\{p(c^*(\theta))\}, & \text{for other case} \end{cases} \quad (19)$$

By convexity of $c_p(\cdot)$, it's easy to know $d\pi/dp = -\alpha G^n(1-p) + p G(1-p)/n + c^*(\theta)$ is concave. Therefore, the derivative of $d\pi/dp$ is non-positive at the point of the greater root. That means the second derivative of expected revenue with respect to p , at the greater root, is non-positive, i.e., $\pi''(p^*(c^*(\theta))) < 0$. It is the second-order sufficient condition for local maximizer. Consequently, we are sure that $p^*(c^*(\theta))$ is the local maximizer, while the smaller root is not.

Since the domain of expected revenue (16) is closed set, in order to find global maximizer, what we should do is to compare the boundary solutions with interior solution. Next, define a $\hat{c}(\theta)$ as follows,

$$\hat{c}(\theta) = \max\{c^*(\theta) \mid \pi'(c^*(\theta)) - \pi(p^*(c^*(\theta)) \mid c^*(\theta)) \geq 0\} \quad (20)$$

and

$$\frac{d\{\pi(0 \mid c^*(\theta)) - \pi(p^*(c^*(\theta)) \mid c^*(\theta))\}}{dc^*(\theta)} = -2L(1-p) \leq 0 \quad (21)$$

It follows that if there exists $\hat{c}(\theta)$, such that

$$\pi(\theta \mid p \in [0, \hat{c}(\theta)]) \geq 0$$

then for all $c^*(\theta) : 0 < c^*(\theta) < \hat{c}(\theta)$ we have

$$\pi(\theta \mid p \in [0, \hat{c}(\theta)]) \geq 0 \quad (22)$$

Essentially, in order to compare boundary solutions with interior solutions, we need to define $\hat{c}(\theta)$. In other words, $\hat{c}(\theta)$ is the *separating point* in which boundary solution is better than interior solution, and the following result is immediate.

Theorem 1. *The following constitutes an equilibrium of our Multi-attributes procurement auction model.*

The simple bidders with type θ_i follow the strategy:

$$\begin{cases} q(\theta_i) = \arg \max [s(q) - c_r(q, \theta_i)], r = 1 \text{ or } 2 \\ p(\theta_i) = c(q, \theta_i) + \int_{\theta_i}^{\bar{\theta}} c(q(t), t) \left[\frac{1 - F(t)}{1 - F(\theta_i)} \right]^n dt \end{cases} \quad (23)$$

The diverse bidder with type θ follows the strategy:

Case 1: For any given $c^*(\theta)$, there exists a unique solution to (17)

$$\begin{cases} q_s(\theta) = \arg \max [s(q) - c(q, \theta)] \\ p(\theta) = p(c^*(\theta)) \end{cases} \quad (24)$$

Case 2: For any given $c^*(\theta)$, there exist two solutions to (17)

$$\begin{cases} q_s(\theta) = \arg \max [s(q) - c(q, \theta)] \\ p(\theta) = \begin{cases} p^*(c^*(\theta)) & \text{if } c^*(\theta) > \hat{c}(\theta) \\ 0 & \text{if } 0 < c^*(\theta) < \hat{c}(\theta) \end{cases} \end{cases} \quad (25)$$

4. CONCLUSION REMARKS

In this paper, we construct a multi-attributes procurement simultaneous auction for 2 objects, through a first-score (price), sealed-bid format. There are two kinds of bidders, *simple bidder* and *diverse bidder*, are considered in our model. Each *simple bidder* is interested in only one of the objects, while the *diverse bidder* is interested in both. With some assumptions, we obtain the equilibrium bidding strategies for the both two kinds of bidders. Basically, the difference between ours work and past correlated research (Krishna and Rosenthal (1996)) are: multi-attributes vs. single attribute (price), first-score(price) vs. second-price, and, most important, procurement auction vs. selling auction. Also, this piece of research can not be realized without Che (1993)'s work about one-unit multi-attributes procurement auction. This work could be regarded as a first step of constructing equilibria and designing mechanism for multi-attributes procurement combinatorial auctions. Future work includes extensions to multiple *diverse bidders* in primal model, and heterogeneous items assumption. It may also be interesting to study collaboration and competition relationship between bidders in multi-attributes procurement combinatorial auctions, in more general or practical environment, which could be undertaken on the basis of this work.

ACKNOWLEDGEMENT

This paper was partly supported by National Science Foundation of China under grant 70231010 and 70321001.

REFERENCES

[1] Che, Y. K., "Design competition through multidimensional auctions", *The Rand Journal of Economics*, Vol.24, pp668-681,1993
 [2] Cantillon, E., Pesendorfer, M., "Combination Bidding in Multi-unit Auctions", *Working paper*, Harvard Univ.,2003
 [3] Krishna, V., Rosenthal, R.W., "Simultaneous Auctions with Synergies", *Games and Economic*

Behavior, Vol. 17, pp1-31, 1996.

[4] Krishna, V., Perry, M., "Efficient Mechanism Design", *Working paper*, Pennsylvania State University, 2000

[5] Nisan, N., "Bidding and allocation in combinatorial auctions", *Working paper*, School of Computer Science and Engineering, Hebrew University of Jerusalem, 1999.

[6] Parkes, D., "Iterative Combinatorial Auctions: Achieving Economic and Computational Efficiency", *Doctoral thesis*, University of Pennsylvania, 2001

[7] Riley, J. G., Samuelson, W. P., "Optimal Auctions", *American Economic Review*, Vol.71, pp381-392, 1981

[8] Rosenthal, R.W., Wang, R. Q., "Simultaneous

Auctions with Synergies and Common Values", *Games and Economic Behavior*, Vol.17,pp32-55, 1996

[9] Rothkopf, M. H., Aleksandar P., Ronald M. Harstad, "Computationally Manageable Combinational Auctions", *Management Science*, Vol. 44, pp1131-1147, 1998

[10] Ron Holzman al. et., "Bundling Equilibrium in Combinatorial Auctions" working paper, Technion-Israel Institute of Technology, 2003

[11] Sandholm, T. W., "Approaches to winner determination in combinatorial auctions", *Decision Support Systems*, Vol.28, pp165-176, 2000

Business Interactions with Web Services

Yucef Baghdadi

Department of Computer Science, Sultan Qaboos University, PO Box 36, PC 123 Al-Khod, Oman
ybaghdadi@squ.edu.om

ABSTRACT

Business interactions require technologies with attractive features to dynamically connecting enterprise information systems and external applications. This work presents a framework of business interactions that consists of: (1) a conceptual architecture view, (2) a technical architecture view, and (3) an implementation architecture view of business interactions. The conceptual view defines and categorizes business interactions into enterprise interactions, partners' interactions and customers' interactions. The categorization allows deciding and deploying the technical view that specifies the required connections for EAI, CRM, B2C and B2B. The implementation view shows how the features of Web services, a connecting technology, enables any kind of specified business interactions. Web services features allow first unlocking and adding value to existing enterprise assets such data and applications; then dynamically connecting them into new solutions such as business processes that respond to business events or changes in business conditions.

Key Words: Business Interactions, Framework, EAI, CRM, B2C, B2B, e-Business, Integration, Web Services, SOA.

1. INTRODUCTION

Businesses need to adapt rapidly and cost-effectively to new businesses events and conditions, namely customers and partners demand and behavior. This requires technologies that enable cost-effective and dynamic interactions within and especially across the boundaries of the enterprise. However, to deploy any technology, one must abstract the business according to different but related perspectives such as in [16]. This work concerns with a business interactions perspective, namely a framework with three views of business interactions in order to decide and deploy a technology with attractive features, namely connecting cost-effectively existing assets in terms of data and applications.

The framework aims mainly at providing business interactions oriented architecture that considers both business perspective and IT perspective in order to point up design as well as technical issues that may hinder the deployment of business interactions enabling technology. That is, Web services.

The framework consists of three architectural views of business interactions perspective that are:

1) The conceptual view architecture aims at specifying the business interactions regardless of the implementing technology. The conceptual view architecture defines, specifies and categorizes business interactions into customers' interactions, enterprise interactions and partners' interactions. This categorization allows a specification of each category of interactions in order to decide an appropriate enabler technology for respectively EAI, CRM, B2C, B2B, and ultimately dynamic e-business.

2) The technical view architecture aims at identifying: (i) the existing enterprise informational as well as computational assets (e.g. existing applications, integration applications and middleware), (ii) the existing (or to develop) connections between the enterprise assets and external applications, and (iii) the requirements and features of the connecting technology. The technical view focuses on technological requirements to fill the business processes breakups that require connections.

3) The implementation view architecture specifies how Web services technology, the *de facto* integration standard [1]; and also the *de facto* Internet standard instance of the services-oriented architecture (SOA) [2][14], presents attractive features in terms of cost-effective connections within and across the enterprise, i.e. between the enterprise and its customers/partners to enable EAI, CRM, B2C, B2B, and ultimately dynamic e-business. Web services technology is, by essence, a connections technology. Indeed, the underlying standards of such a technology, namely, XML, WSDL, UDDI, WS-I, BPEL4WS, SOAP and other related to security and transactions when matured will allow: (1) interfacing the enterprise assets that are data and applications as services in order to unlock them, (2) publishing the interfaces in a registry to be discovered and reused, (3) discovering the services, (4) connecting (statically or dynamically) any application to the services, (5) ultimately connecting services to each other with respect to SOA architecture, and (6) composing dynamically the services into new solution to respond to business events or changes.

2. CATEGORIZATION OF INTERACTIONS

Due to organizational and technical constraints, business processes are broken up into causal and reciprocal activities. That is, the set of activities form a

flow where each activity takes an input from a provider and serves an output to a consumer. These business processes breakups require a set of business interactions, which are generally implemented by organizational and technological artifacts. Therefore, one needs to abstract these interactions in order to first specify them with respect to a business perspective before to come to a technology perspective. The abstraction allows deciding an appropriate technology that enable cost-effective and dynamic interactions.

2.1 Business Interactions

The concept of interactions has been specified by different disciplines. A business perspective specification, we adapt from different disciplines, considers interactions as a set of reciprocal (or causal) activities performed by local or remote providers and consumers in a synchronous or asynchronous manner. Each of them has a certain degree of autonomy and freedom (e.g. to leave at any time). Therefore, business interactions are characterized by space, time, and dynamics. Moreover, providers and consumers interact in different situations.

Space. Interactions involve local as well as distributed and remote providers and consumers.

Time. Action-reactions of the providers and consumers are temporal, synchronous as well as asynchronous.

Dynamics. While interactions in goal-driven businesses are often rigid and fixed (e.g. mechanisms of coordination used to offset planned task allocation in goal-driven businesses). Interactions, in event-driven businesses, are dynamic relationships among autonomous providers and consumers.

Situations. Situations of interactions are answer to 'why do providers and consumers interact?' With respect to a business (and information systems) perspective. Business providers and consumers interact for numerous reasons, namely:

- Internal and external exchanges of information.
- Synergy of business processes.
- Unlocking informational (e.g. data) and computational resources (e.g. applications).
- Emergent knowledge. Indeed, knowledge resulting from interactions is more relevant and more complete than simple knowledge integration.

2.2 Categorization of Business Interactions

In addition to their specification with respect to space, time, dynamics, and situations, it is a must to categorize business interactions while deciding and deploying enabler technologies. First, business interactions differ according to the types of the involved providers/consumers, and the types of business processes and business events. Indeed, local

business interactions are different from interactions with partners and customers. Moreover, the business interactions of a goal-driven business differ from those of an event-driven business. Second, business interactions will not have the same priority with respect to the enabling technology. For, instance, EDI technology may be suitable for a long term fixed relationship between a business and its partners. Whereas Web services technology allows a real dynamic relationship between a business and its partners. Third, interactions between a business and its partners/customers are based on the effectiveness of internal interactions. That is, one cannot design and implement external interactions if the local ones are not working properly. For instance, B2B cannot be approached if EAI is not working properly. Figure 1 distinguishes each of the following categories of business interactions:

- 1) Interactions between business and its customers (**C-B**).
- 2) Internal (enterprise) interactions that are:
 - Interactions between enterprise (local) primary processes (**P-P**).
 - Interactions between local primary and supporting processes (**P-S**).
 - Interactions between local business process activities (**AT**). These are required when a local business process is decomposed (**A-A**).
 - Interactions between business processes and the business objects (or coordination artifacts) as presented in the different information systems and legacy systems.
- 3) Interactions across businesses, i.e. interactions between business processes that cross the boundaries of the organization (**B-P**).

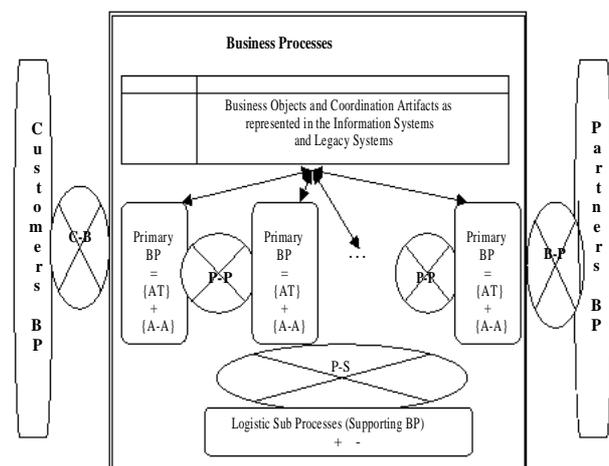


Fig. 1. Categories of Business Interactions

2.3 Interactions Common Activities

Regardless of their characteristics, i.e. dynamics, time, space, situations and categories, interactions require announced partners to perform the following activities as shown in Figure 2: (a) Identify the partner, (b) Send a message to the partner, (c) Read and interpret the message, (d) Understand the message, (e) Act

according to the interpretation and understanding (semantics required), (f) Receive a message, (g) Read and interpret the message, (h) Understand the message, and (i) Re-act according to the interpretation and understanding (semantics required).

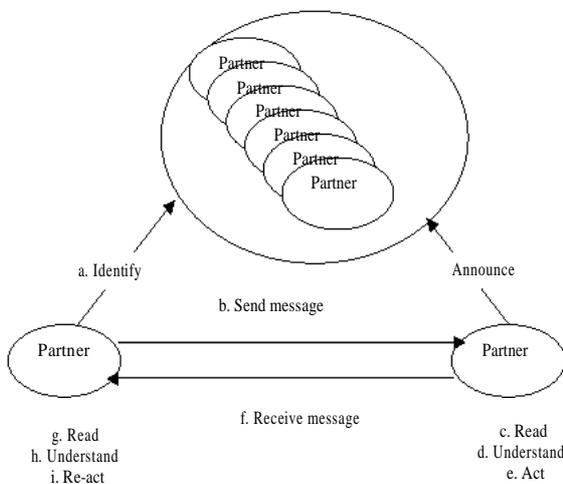


Fig. 2. Interactions Common Activities

3. TECHNICAL ARCHITECTURE VIEW

The technical architecture view (Fig. 3) aims mainly at specifying the whole business IT system as a set of connecting IT subsystems that support business processes. The IT subsystems may be any combination of information systems, legacy systems and other kinds of applications. The interacting IT subsystems are:

1. Local subsystems that are different enterprise information systems (EIS) and legacy systems (LS) used to sustain primary and supporting business processes.
2. Clients subsystems that interact with local IT subsystems, these are stand-alone clients (SC: clients that interact through standalone applications), Web browsers (BC: clients that interact through browsers) and mobile clients (MC: clients that interact through mobiles).
3. Partners subsystems that interact with local IT subsystems, these are the partners information systems, legacy systems, components, objects or services.

The connecting IT subsystems are interfaced and viewed as endpoints. That is, their internal behavior is encapsulated as a set of endpoints. The interfaces depend on the types of IT subsystems. That is, legacy systems and to a less extent enterprise information systems or applications will be wrapped to expose public interface. While, the components, objects and services are, by essence, specified as interfaces and implementations. The interfaces are:

1. Interfaces of the IT subsystems that are implemented with enterprise applications integration (EAI)

according to the integration orientation (data, applications or business processes) and different types of middleware (transactional, remote procedure call, messaging, message brokers, distributed objects computing, or SOAP).

2. Interfaces of the partners IT subsystems that are implemented through B2B electronic commerce applications (e.g. EDI, ebXML, Web services) or e-market brokers.
3. Interfaces of the clients that are implemented through B2C and CRM applications.

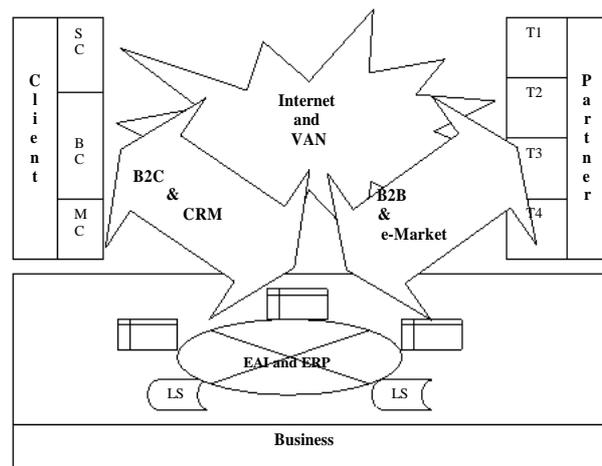


Fig. 3. Technical View of Business Interactions

This connection-oriented technical architecture view represents different categories of business interactions. To effectively allow dynamic business transaction, the implementing technology must emphasize connection of transparent and self-contained existing enterprise assets that don't tightly depend on each other. That is, a technology that allows loose coupling of the interacting elements. Web services technology is the *de facto* connections technology when deployed with respect to SOA architecture.

4. IMPLEMENTATION WITH WEB SERVICES

The technical architecture view is implemented with different technologies. Indeed, business interactions have been implemented as EAI, ERP, B2C, CRM, and B2B using traditional middleware, and distributed object-computing middleware (e.g. CORBA, DCOM and RMI) before the advent of SOA architecture and its instances, namely e-services and Web services. SOA architecture with Web services technology, despite the lack of maturity of the underlying standards, presents advantages with regard to (1) loose coupling of interacting elements, (2) dynamics, and especially (3) the connection-oriented applications of Web services.

4.1 SOA Architecture

SOA architecture was first described by Gartner Group in 1996 as architecture based on the concept of service. It is essentially a collection of services that communicate with each other [2]. There exist several definitions of the term service. For instances, a kind of business oriented definition is given by [10] who defines a service as “business function of an application, so that another application or an application at another enterprise may find it useful to invoke”. A more technical definition is given by [13] who defines a service as “a location-transparent, network-addressable, invocable unit of software logic that is well defined, self-contained, and does not depend on the context or state of other services”.

We consider a service as a self-contained element of the information system that implements business logic, business rule, or data operation (insert, update, delete, or retrieve data); and provided with a well-defined and standardized interface to be transparently accessible. This element of the information system must be designed and deployed with respect to SOA architecture. SOA aims mainly at achieving communication among loose coupling interacting software agents [8]. Therefore, Web services technology is the most attractive due to its underlying standards that enable communication between services within and outside the enterprise through a public network that is Internet.

4.2 Web Services as Internet Instance of SOA

W3C/WS Architecture Group defines Web services as “software system identified by URI, whose public interface and bindings are defined and described using XML. Other software can discover its definition. These software may then interact in a manner described by its definition using XML-based messages conveyed by Internet protocols” [1].

Web services technology is mainly a collection of technologies and standards that allow connecting services over the Web. These technologies and standards allow interfacing, publishing, binding, composing services through communications protocols based on Internet protocols (e.g. HTTP, SMTP, MIME). The standards XML, WSDL, UDDI, WS-I, BPEL, and SOAP allow communication and execution via the Web of self-contained and loosely coupled services within and outside the enterprise, which makes Web services technology not only the de facto integration standard, but also the de facto Internet standard instance of SOA architecture [14]. Web services technology can easily live with other technologies such as CORBA, DCOM, or RMI. These features of Web services will enable all the categories of business interactions as shown in the next section.

4.3 Web Services Enabler of Business Interactions

The above interactions common activities can be easily implemented by the standards and technologies underlying Web services. Figure 4 shows how Web services stack and technologies (bolded in the figure) implement the interaction common activities.

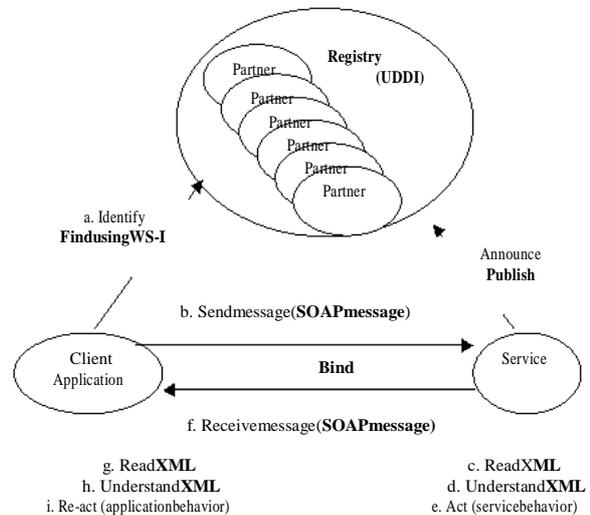


Fig. 4. Web Services and Common Activities

4.4 Web Services Enabler of EAI, B2B, CRM, B2C, and Dynamic e-Business

The features of Web services technology, notably their communication features enable all the categories of business interactions as summarized in Table 1.

4.4.1 Web Services Enabler of EAI

EAI is an integration technology. It consists of providing interface to technologies implementing heterogeneous enterprise information systems and legacy systems. It aims at making applications composite to rapidly respond to business events, and to adapt to change in business conditions. The main features of EAI are:

- Connection of the applications (in different enterprise information systems and legacy systems).
- Introspection to look up and find applications or databases in order to connect them.
- Translation of data and messages exchanged between the applications and databases of the connected systems
- Control of flow of the applications, from within the different connected systems, involved in the composition of business processes.

These features can easily be enhanced with Web services technology. Indeed, Web services technology present transparent view of business logic, rules and data by interfacing and publishing them to be accessed, invoked, and consumed by applications that enter in the composition of primary and supporting business process.

EAI will easily support Web services. Indeed, Web

services are just a step in the evolution of technology and standards. The standards underlying Web services will ultimately simplify application integration within and outside organizational boundaries. These standards (when matured enough) will allow Web services to become the de facto integration standard by supporting all features of EAI. Moreover, Web services reduce the complexity and the cost of the traditional integration middleware.

4.4.2 Web Services Enabler of B2B

A B2B application consists of connecting applications from different businesses to exchange agreed upon (or standardized) business documents (e.g. order, invoice). For large organizations, this has been achieved based on EDI specifications. Despite the evolution of EDI, from EDI data transmission, VAN (used to simplify complexity of multiple connections), DCOM and CORBA (used to reduce the cost related to VAN and to standardize the message), to XML as EDI standards, there are still challenges that can be addressed by XML and Web services. Indeed, Web services (with matured underlying technologies and standards) are used in B2B to:

- Reduce the cost of entry into B2B for small and medium businesses. Indeed, EDI deployment and maintenance is very costly, which deprives small and to a less extent medium businesses to play the right role they are intended to in the economics arena.
- Allow effective dynamics. Indeed, EDI specifications assume a long term fixed and well-specified interactions (e.g. agreed upon format) deployed through proprietary networks. Web services connecting technology is based on the contract of type "take or leave", which gives more freedom and dynamics to businesses to choose transparent services accessed through the protocols of the Internet (a public network).
- Fix the problem of the difference in semantics, and the problem of fixed record format of exchanged between businesses applications as imposed by EDI specifications.

4.4.3 Web Services Enabler of CRM

CRM system is an information system that records information such as customer contact information. A CRM system is generally used by customers and representatives. A CRM system is not necessary an internal system. On the contrary, external CRM systems are better suited to be accessed as Web services. Indeed, CRM as Web services will certainly present better services, performance and reliability.

4.4.4 Web Services Enabler of B2C

Web services will facilitate the development and deployment of applications that can be readily accessed by PDA, office devices, and mobile devices in addition

to Web browsers. Web services will permit applications developers to better leveraging existing businesses logic, rules and data by invoking them by any kind of clients. Moreover, Web services will reduce search complexity for exiting services.

4.4.5 Web Services Add-Value

Web services features add value to traditional applications used to implement internal and external business interactions as summarized in Table 1.

Table 1. Value added by Web services

Interactions Applications	Value added by Web Services Technology
EAI	<ul style="list-style-type: none"> • Standardized interfaces of the elements of the enterprise information systems and legacy systems. • Connection of loosely coupled elements of the enterprise information systems and legacy systems. • Introspections through UDDI and WSDL. • Use of XML to exchange and translate messages. • Control of the flow of the integrated applications through BPEL.
B2B	<ul style="list-style-type: none"> • Entry into B2B to small and medium enterprise (VAN is no longer required). • Dynamic business interactions. Businesses are really autonomous in their interactions • Free message format using XML. Messages format not fixed and standardized.
CRM	<ul style="list-style-type: none"> • Use of external as well as internal CRM applications accessible from anywhere through Internet.
B2C	<ul style="list-style-type: none"> • Applications involved in B2C are better interfaced (WSDL) to be accessible through different means, i.e. office devices, PDA and mobile devices in addition to Web browsers. • Search and tracking facilities exposed as Web services.
Dynamic e-Business	<ul style="list-style-type: none"> • Dynamic binding facility allows business applications or services to bind to services at run-time. • UDDI and WS-I allows inspection and selection of efficient services and cost-effective connection.

5. RELATED WORK

Business interactions have been implicitly seen as an integration problem with an IT perspective rather than a business perspective. That is, the issue is often how to make information systems (including applications and databases), running on different platforms, interoperable to exchange structured as well as unstructured data? Various integration ways have been approached, namely: (1) a data-oriented integration (e.g. distributed databases [3]), (2) an object-oriented integration based on distributed object computing middleware (e.g. COBRA, DCOM, RMI), and (3) a

service-oriented integration based on SOAP middleware (e.g. e-services, Web service).

The object-oriented integration approach makes applications interoperable through a broker used for discovery and invocation. In this category, CORBA, DCOM and RMI were designed mostly from an IT perspective in order to help developers operate more effectively [15]. Due to their great returns and quality productivity, these technologies have been used in EAI, and B2B to offset the complexity related to value added networks (VAN) used as brokers in EDI. However, they do not allow a loose coupling because they are built on their own technology.

In the last years, clear trend is to move away from data-oriented and object-oriented integration to service-oriented integration [2][5][7][8][9][15].

The state of art shows SOA and Web services with praise. This technology is now widely adopted because it allows connecting partners (e.g. B2B integration) with reduced cost [4]. The ultimate goal of this technology is to enable dynamic e-business. Web services technology is a main focus of various software development companies [6]. It is considered as the hottest topic by software industry.

Our approach presents a model of business interactions as a framework that allows guidance towards a method to deploy Web services as add value for the existing applications that are EAI, B2C, CRM, and B2B. The framework is used to abstract and then deploy a comprehensive and multipurpose set of services with respect to SOA. A model of business interactions simplifies the vision and the technological architecture.

6. CONCLUSION

We have proposed a framework for business interactions, as one of the most important perspectives of a business modeling besides the business processes, businesses objects and business events perspective. An interactions perspective is critical while deciding the alignment and deployment of an enabling connecting technology. This abstraction is made up of three architectural views of business interactions: conceptual, technical, and Web services.

The main goal of the business interactions abstraction is to propose recommendations and guidance that help generating, and deploying a set of Web services with respect to SAO architecture. The abstraction makes clear the breakups of the business processes and the technological requirements to fill these breakups.

We have shown how the features of Web services technology as Internet instance of SOA architecture add value to the features of EAI, B2C, CRM, B2B in order to ultimately allow dynamic e-business.

Deploying Web services technology with respect to architecture is a critical issue nowadays where businesses need to cost-effectively and dynamically integrate business processes that cross their boundaries, which is critical for their survival.

This work can be extended by developing models and supporting tools towards a method for deploying Web services as an Internet standard instance of SOA.

REFERENCES

- [1] Austin D et al., "Web Services Architecture Requirements", W3C Working Group Draft 14, Retrieved from <http://www.w3.org/TR/wsa-req#id2604831>, 2002.
- [2] Barry D K. , "Web services and Service-Oriented Architecture", Elsevier (USA). 2003.
- [3] Batini C, Lenzirini M, Navathe S. B., "A Comprehensive Analysis of Methodologies for Database Schemas", *Integration. ACM Computing Surveys*, 18(4): pp. 322-364, 1986.
- [4] Box B., "Web Services Revolution", Retrieved http://www.eds.com/thought/thought_leadership_web_revolution.pdf, 2003.
- [5] Cohen L. D., "Building and Managing Adaptive e-Business Solution Infrastructure", Keynote in proceedings of IEEE CEC03, 2003.
- [6] Fensel D, Bussler C., "The Web Service Modeling Framework WSMF", *Electronic Commerce Research and Applications* 1(1): pp. 113-137, 2002.
- [7] Frankel D, Parodi J., "Using Model-Driven Architecture to Develop Web Services", Retrieved <http://www.omg.org/attachments/pdf/WSMDA.pdf>, 2002.
- [8] He H., "What is Service-Oriented Architecture?", In: O'Reilly Open Source Convention: July, pp. 26-30, 2003.
- [9] Linthicum D. S., "Moving to Service-Based Application Integration", *eAI Journal*, October: pp. 9-10, 2001.
- [10] Natis Y., "Service Oriented Architecture Ushers in the Next Era in Business Software Engineering", *BIJ*, May: pp. 23-25, 2004.
- [11] Plesko D and Lee J. M., "Web Services and Enterprise Integrations: Friends, Not Foes", *BIJ*, October, pp. 8-12, 2003.
- [12] Phillips J and Foody D., "Building Foundation for Web Services", *eAI journal*, March, pp. 8-13, 2002.
- [13] Roth P. , "To a Service-Based Architecture", *BIJ*, October: pp. 48-50, 2003.
- [14] Smith D. Web Services Enable Service-Oriented and Event-Driven Architecture, *BIJ*, May: pp. 12-14, 2004.
- [15] Wong L. , "Services: A Key Component for Success", *eAI Journal*, March: pp. 18-25, 2001.
- [16] Zackman J. A., "Concepts of the Framework for Enterprise Architecture", Zackman International Los Angeles, CA, 1996.

Collaboration Strategy of Business-to-Business Enhancing Customer Satisfaction

Chulsoo Kim¹, Yon Tae Kim²

¹ Associate Professor, College of Business, Inha University
253, Yonghyun-dong, Nam-gu, Incheon, 136-701, South Korea, cskim@inha.ac.kr

² Lectures, Undergraduate College of Business, Inha University
South Korea, bsicall@nate.com

ABSTRACT

This paper focuses on demand chain management, which accomplishes the synchronization of demand with supply through collaboration of buyer and supplier across the chain. And the purpose of it is to find the relationship between four integration strategies produced by combining supply integration and demand integration and its impacts on two performances, operating efficiency and customer satisfaction. The relationship is analyzed with respect to manufacturing and services respectively. Interestingly, our results show that the strength of integration affects differently two performances.

Keywords: Supply chain; Web-based Integration Strategy; Demand integration; Supply integration; Efficiency; Customer satisfaction

1. INTRODUCTION

A supply chain is composed of components involved in designed to produce products and services, procuring raw materials, transforming them into semifinished and finished products, and delivering them to end customer. In contrast, demand chain management (DCM) is defined a practice that manages and coordinates the whole supply chain, starting from the end customer and working backwards to raw material suppliers (Vollmann et al., 2000). For more specific, DCM is focusing on synchronizing demand with supply through customer and partner collaboration across multiple channels, while supply chain management (SCM) is managing and optimizing the sourcing, production and distribution operations of the enterprise to match a given demand (Woods, 2002).

Why Demand Chain Management is important? It is because three market dynamics are increasing the complexity and criticality of intelligently executing upon supply and demand chain processes: increased demand volatility, accelerating product changes and highly distributes supply. It is virtually impossible to 100% accurately forecast demand. Today the volatility of market demand for most products has increased. The costs of inaccurate forecasts have increased due to smaller profit margins, more rapid product price drops, and shorter product lifecycles. Demand volatility has also increased due to the expansion of distribution and sales channels. Companies continue to innovative and develop more advanced and sophisticated products that serve consumers' and businesses' appetite for technology and its uses. The supply-side network for many manufacturing companies has experienced fundamental changes. Companies have increased their manufacturing outsourcing and have become more reliant upon a broader network of component and customer suppliers.

Demand volatility, product innovation and distributed supply networks are creating new business challenges for manufacturing companies and their supply and demand chain partners. Information latency and lack of visibility across this chain of partners result in a situation where the manufacturing companies can not effectively manage the new market volatility.

To be efficient and contribute to high customer satisfaction is one of important challenges in manufacturers and services. Through e-business, it is possible to serve customers individually with customized bundles of goods and services. However, it is difficult to improve both operational efficiency and customer satisfaction. Implementing a rigid approach in management would lose customer satisfaction. On the other hand, getting customization would decrease efficiency. For a supply chain, demand-side companies enhance high customer satisfaction composed of fill rate, customer's perception of the service from them, etc., whilst supply-side ones try to increase operational efficiency composed of lead-time, inventory level, inventory turnover, etc.

In order to achieve higher operational efficiency and customer satisfaction, DCM should be extensive up- and downstream integration between companies involved in the chain. Throughout the Internet, the integration has only become possible nowadays. And the tradeoffs between low cost, speedy transaction, diverse service, broad networks and rich content across the chain are resolved effectively through the web.

To be successful implementation of DCM, it is necessary that DCM has three components: Integrate, Exchange and Synchronize (Woods, 2002). Firstly, Integrate is performing data integration over the chain; secondly, Exchange is enabling collaboration with trading partners

to share intellectual capital about demand. Finally, Synchronize is creating business processes, not just data synchronization, that integrates knowledge about the chain performances to shape demand in this extended value network. Also, web-based technologies are increasingly indispensable for not only exchange and integration but also knowledge synchronizing.

Korea firms are mostly on the level of high position in infrastructure of the Internet. But it is not for e-business or business process. It correctly says that they are now struggling in integration of the Web with their business processes. So we sampled the companies posted in higher level on e-business processes.

By using supply-side or demand-side integration across the chain, demand and supply integration strategies address four strategic alternatives (Frohlich and Westbrook, 2002). They are shown in Fig. 1. Whilst model A indicates a strategy of little or no web-based integration of both sides, model D is oppositely a strategy with high levels of web-based integration collaborating with both sides partners. In between the other dimensions are strategies with supply-side integration of companies (model B) and demand-side integrations (model C). Adopting a web-based integration strategy requires collaborative agreements of problems on sharing information, benefit, knowledge of industry and market, etc. with suppliers and customers. Throughout the agreements, a company could decide a long-term plan for implementing strategies. To generate the plan effectively, a company should investigate how much the performances, operational efficiency or customer satisfaction, are increased through adoption of each strategy.

The objective of our study is to find the relationship between adopting web-based integration strategy and its impacts on two performances with manufacturing and services in Korea. For more specific, this research investigates whether there is a difference between the impacts of each selected strategy on two performances: operational efficiency and customer satisfaction, and also a difference between the impacts in manufactures and services.

2. LITERATURE REVIEW

Substantial previous research highlights many opportunities separately associated with either supply and demand integration in supply chain (Table 1). As Handfield (1993) mentioned, supply integration means "obtaining frequent deliveries in small lots, using single or dual sources of supply, evaluating alternative sources on the basis of quality and delivery instead of price, establishing long term contracts with suppliers, reducing buffer inventories and eliminating formal paperwork." The supply integration does not only reduce costs but also improve lead times. Through the Internet, other benefits include improved supplier reliability (Carr and

Peason, 1999) and communication (Freeland, 1991).

In terms of the demand-side of companies, major challenges are improving demand visibility and planning in supply chains (Fisher et al., 1994). As Lee et al. (1997) noted, distorted demand information in a supply chain would incur tremendous inefficiencies: excess inventories, poor customer service, lost profits, misguided capacity plans, ineffective transportation and missed production schedules. One trend is integrating up- and downstream information to coordinate non-vertically integrated firms (Marbert and Venkataramanan, 1998). Due to magnification of unstable planning over supply chains, it needs controlling this amplification, which is essential to good demand management. Information technologies involving the web would make the demand integration successfully, which could include shared data between planning and control systems (Bowersox and Daugherty, 1995).

In demand chains, other interested issues are improving delivery efficiently and rapidly (Bhatnagar and Viswanathan, 2000; Cachon and Fisher, 2001; Heikkila, 2002) and logistics communication (Corbett et al., 1999). Basics of the core of demand-driven supply chain integration strategy are demand forecast integration, inventory reductions and the elimination of non-valued added activities.

E-business— the use of Internet-based computing and communications to execute both front-end and back-end business processes — has emerged as a key enabler to drive supply chain integration (Lee and Whang, 2001). The Internet creates a new environment for exchanging information and conducting business transactions. The most direct effect of the Internet is to create new opportunities to improve the efficiency and effectiveness of the operation of the supply chain. This is because of the cost-effective capacity to generate visibility across all aspects of the supply chain, including point-of-sale information, manufacturing schedules, vendor stocks, customer inventories, demand patterns, sales/marketing initiatives, and carrier schedules.

And the Internet is an enabler to actually combine customers and suppliers over a supply chain. Pre-internet, real-time demand information and inventory visibility were almost impossible to achieve and most supply and demand integration involved a patchwork of telephoning, faxing, and EDI. Web-based technologies allow customers and suppliers to co-operate each other for inventory planning, demand forecasting, order scheduling, target marketing, and customer relationship management. And the greatest potential of the Internet is being realized by speeding up communication between customers and their suppliers, improving service levels, and reducing logistics costs (Lancioni et al., 2000). Lee and Whang (2001) have offered a view emphasizing e-business applications to the supply chain: planning and

execution of the front-end and back-end operations in a supply chain using the Internet.

Throughout the Internet, traditional demand and supply integration practices present four strategic alternatives (Frohlich and Westbrook, 2002). The worst is a strategy of little or no web-based integration (model A). Next, two models are introduced groups whose strategies involve web-based integration with either their suppliers (model B) or customers (model C). And last is a strategy with high levels of web-based integration, which requires being cooperative among the whole demand chain members from customers backwards to suppliers called DCM (model D). Frohlich and Westbrook(2002) suggests that the higher the level of integration intensity with suppliers and customers, the greater the benefits. Our study involves above four strategies.

Through the Web, firms are able to obtain global visibility across their extended network of trading partners and respond rapidly to diverse customer demand. Applying the Web into their demand chain makes them more beneficial to efficiency in internal view and customer satisfaction in external view.

Web-based DCM strategy performs the highest operational efficiencies; faster delivery time, reduced transaction costs, greater profitability and enhanced inventory turnover (Frohlich and Westbrook, 2002). And customer satisfaction is composed of higher fill rate, customer loyalty, recommendations, and enhancing customer's perception on pricing and the quality of a product. Heikkila (2002) studied on a demand chain of mobile telecommunication technology capable of increasing operational efficiency and customer satisfaction. In the chain, high customer satisfaction is more important rather than efficiency. The problem of focusing one of both would depend on a company's situation: type of its role as a component in the chain, type of a product, etc.

To high operational efficiency or customer satisfaction, DCM is required extensive up- and downstream integration between companies involved in the chain. Throughout the web, the integration could be very speedy, and also, it can effectively resolve the tradeoffs between low cost, speedy transaction, diverse service, broad networks and rich content across the chain.

3. RESEARCH METHOD

3.1 Research model and hypotheses

As Narasimhan and Jayaram (1998) studied, a low integration strategy (model A) in the Internet era should deliver the fewest benefits among four integration strategies. The DCM strategy (model D) should deliver the highest levels of performance. The higher the level of traditional integration with suppliers and customers the greater the benefits, as Frohlich and Westbrook (2001)

noted, and the Internet should only reinforce this relationship. The performance for the supply integration (model B) and the demand integration (model C) strategies should logically fall between the low integration (model A) and DCM (model D) approaches. Fig. 1 shows the research model and study hypotheses.

The following arguments lead to the set of hypotheses.

H1a. Manufacturers and services adopting a web-based DCM integration strategy (model D) will have the highest levels of operational efficiency.

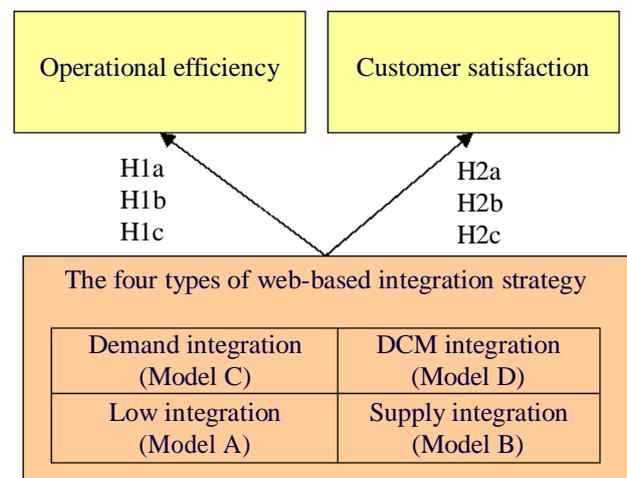
H1b. Manufacturers and services adopting a web-based supply (model B) or demand (model C) integration strategy will have medium levels of operational efficiency.

H1c. Manufacturers and services adopting a low web-based integration strategy (model A) will have the lowest levels of operational efficiency.

H2a. Manufacturers and services adopting a web-based DCM integration strategy (model D) will have the highest levels of customer satisfaction.

H2b. Manufacturers and services adopting a web-based supply (model B) or demand (model C) integration strategy will have medium levels of customer satisfaction.

H2c. Manufacturers and services adopting a low web-based integration strategy (model A) will have the



lowest levels of customer satisfaction.

Fig. 1. Research model and hypotheses

3.2 Operational definitions and measure items of variable

Some factors mentioned in this research are extracted through existing studies of Table 1. The others, especially related to customer satisfaction factor, are through the relevance of demand chain factor. Operational definitions, measure items of variable and sources of variables are described in Table 1.

As it mentioned in Section 2, it generates four integration strategies by using the two types of integration. While the factor of operational efficiency describes company's

performance: delivery time, transaction cost, profitability, and inventory turnover, through the web-based integration, the factor of customer satisfaction shows customer's perceptions for a company: information support for products and services, intent to continuous

transaction, satisfaction of price, quality and delivery, and likelihood to recommend supplier to other customers. The scale for all measure items depicted in Table 1 is a 7-point Likert scale with 1 indicating not agreed at all, 4 indicating so-so and 7 indicating fully agreed.

Table 1. Operational definitions and measure items of variable

Factor	Operational definition	Measure item	Studies
Supply integration	Levels of web-based supply chain integration with suppliers	Integrated inventory planning with suppliers	Frohlich and Westbrook(2001), Lancioni et al. (2000), Lee and Whang(2001), Rahman(2003)
		Integrated supply chain management with suppliers	Frohlich and Westbrook(2001), Lancioni et al.(2000), Lee and Whang(2001), Rahman(2003)
		Integrated demand forecasting with suppliers	Frohlich and Westbrook(2001), Lancioni et al. (2000), Lee and Whang(2001), Rahman(2003)
		Integrated order scheduling and tracking with supplier	Frohlich and Westbrook(2001), Lancioni et al. (2000), Lee and Whang(2001), Rahman(2003)
Demand integration	Levels of web-based supply chain integration with customers	Targeted marketing/customer profiling	Heikkila(2002)
		Online order taking/receipt	Lee and Whang(2001)
		Integrated demand forecasting with customers	Frohlich and Westbrook(2001), Heikkila(2002), Lancioni et al.(2000), Lee and Whang(2001), Rahman(2003)
		Customer relationship management	Heikkila(2002), Lee and Whang(2001)
Operational efficiency	Levels of achievement by web-based integration with suppliers or/and customers	Faster delivery time	Frohlich and Westbrook(2001), Heikkila(2002), Lancioni et al. (2000), Lee and Whang(2001), Rahman(2003)
		Reduced transaction costs	Lancioni et al. (2000), Lee and Whang(2001)
		Improved greater profitability	Chien et al.(2003), Frohlich and Westbrook(2001), Lee and Whang(2001)
		Enhanced inventory turnover	Frohlich and Westbrook(2001), Heikkila(2002), Lancioni et al. (2000), Lee and Whang(2001)
Customer satisfaction	Levels of satisfaction for purchasing producer's products or services	Information support for products/services	Chien et al.(2003), Frohlich and Westbrook(2001), Heikkila(2002), Lee and Whang(2001), Lancioni et al. (2000)
		Intending to continuous transaction (customer royalty)	Chien et al.(2003), Heikkila(2002)
		Satisfaction of price, quality and delivery	Chien et al.(2003), Frohlich and Westbrook(2001)
		Recommending producer to other customers	Chien et al.(2003)

3.3 Data collection

The survey was developed in three stages. In the first stage, we identified relevant measures of web-based demand and supply integration, operational efficiency, and customer satisfaction in the literature and drafted the instrument. In the second stage, we pre-tested the survey with 90 manufacturers and service firms to further gauge its content validity and overall readability. In the final stage, we executed the survey.

Data were collected from random samples of manufacturing and services from across Korea. Questionnaires were divided of two types. One is for company; the other is for his customer. We sampled from 7 regions of Korea including Seoul and Busan and, in the external validity, the companies sampled in Korea do

well in e-business and the infrastructure of the Internet of Korea is ranked higher. 640 questionnaires are sent to company of manufacturing or services, and his customer by e-mail for convenience. And company's respondents were mostly managers with enough seniority to know about their companies' up- and downstream integration, and operational efficiency. Customers responded the level of customer satisfaction about his supplier. In our study, to gather more accurate response with consideration of Korean characteristic, all of variables in Table 2 were measured by a 7-point Likert scale: 7 = fully agreed, 4 = middle, and 1 = not at all. Returned usable questionnaires were totally 209 and the survey response rate was 32.7%. Responses of company are 118 cases; 40 manufacturers, and 78 services. And responses of customer are 91; 29 manufacturers and 62 services.

4. HYPOTHESES TEST AND RESULTS

This study analyzed the differences, concerning relationships between web-based integration strategies and performance in manufacturing and services, using ANOVA and the Scheffé method. Table 2 shows the results. In the table, demand chain integration strategy (model D) was archived the highest performance, and then demand integration strategy (model C) and supply integration strategy (model B) were next. Low integration strategy (model A) was the lowest.

And in Table 2, the manufacturers that employed demand chain integration strategy (model D) have archived the highest operational efficiency than the other strategies. While there is significant in between adopting model C

and model D, they were not significantly different from other strategies. The manufacturers with model B had higher operational efficiency than model A and lower operational efficiency than model D. In services, model B, model C and model D had not been significantly different each other, but they were significantly different from model A. Using the results of ANOVA, we tested the set of hypotheses. Table 3 represents the results for hypothesis test by manufacturers and services. As a further evaluation of H1a, H1b and H1c for operational efficiency, and H2a, H2b and H2c for customer satisfaction, the total of manufacturers and services were merged together and regression was run to test effect of web-based integration (models A-D) on performance for operational efficiency and customer satisfaction separately.

Table 2. ANOVA for performance by integration strategy and company type

Cluster	Manufacturers ^a						Services ^a						Total ^a					
	Operational efficiency			Customer satisfaction			Operational efficiency			Customer satisfaction			Operational efficiency			Customer satisfaction		
	m	s	n	m	s	n	m	s	n	m	s	n	m	s	n	m	s	n
Model A	2.40	1.19	5	4.13	1.59	2	3.13	1.65	14	4.71	0.93	8	2.93	1.54	19	4.60	1.01	10
	(B*, D**)			-			(B***, C**, D***)			-			(B***, C**, D***)			-		
Model B	3.78	1.01	8	5.18	0.66	7	4.74	1.23	18	4.88	1.01	14	4.44	1.24	26	4.98	0.90	21
	(A*, D***)			-			(A***)			-			(A***, D*)			-		
Model C	3.50	0.40	5	4.90	0.72	5	4.65	1.39	17	5.13	0.71	15	4.39	1.32	22	5.08	0.70	20
	(D**)			-			(A**)			-			(A**, D*)			-		
Model D	5.28	0.85	22	5.40	1.01	15	5.16	1.06	29	5.45	1.11	25	5.21	0.97	51	5.43	1.06	40
	(A***, B***, C***)			-			(A***)			-			(A***, B*, C*)			-		
Total	4.40	1.38	40	5.17	0.94	29	4.58	1.46	78	5.14	1.00	62	4.52	1.43	118	5.16	0.97	91

a. m: mean; s: standard deviation; n: number of cases; alphabets in parentheses show which other integration strategy type the cluster's means performance is significant different from.

*: P < 0.10, **: P < 0.05, and ***: P < 0.01 significance.

Table 3. Results of hypothesis test

Hypothesis		Decision
H1a	Manufacturers adopting a web-based DCM integration strategy (model D) will have the highest levels of operational efficiency.	Supported
	Services adopting a web-based DCM integration strategy (model D) will have the highest levels of operational efficiency.	Not supported
H1b	Manufacturers adopting a web-based supply (model B) or demand (model C) integration strategy will have medium levels of operational efficiency.	Partial supported (only Model B)
	Services adopting a web-based supply (model B) or demand (model C) integration strategy will have medium levels of operational efficiency.	Not supported
H1c	Manufacturers adopting a low web-based integration strategy (model A) will have the lowest levels of operational efficiency.	Partial supported
	Services adopting a low web-based integration strategy (model A) will have the lowest levels of operational efficiency.	Supported
H2a	Manufacturers adopting a web-based DCM integration strategy (model D) will have the highest levels of customer satisfaction.	Not supported
	Services adopting a web-based DCM integration strategy (model D) will have the highest levels of customer satisfaction.	Not supported
H2b	Manufacturers adopting a web-based supply (model B) or demand (model C) integration strategy will have medium levels of customer satisfaction.	Not supported
	Services adopting a web-based supply (model B) or demand (model C) integration strategy will have medium levels of customer satisfaction.	Not supported
H2c	Manufacturers adopting a low web-based integration strategy (model A) will have the lowest levels of customer satisfaction.	Not supported
	Services adopting a low web-based integration strategy (model A) will have the lowest levels of customer satisfaction.	Not supported

5. CONCLUDING REMARKS

About impacts of four strategies on two performances, it is not significant at level of 95% that there are differences of the impacts on customer satisfaction by four strategies regardless of industry, while it is significant in the difference of the impacts on efficiency by manufacturers and services. Through the mean values on performance of efficiency, the model D, for manufacturing companies, shows the highest value and next orders are B, C, and A. The result means that a manufacturing company focuses on supply integration of suppliers and thereby accomplishing the internal efficiency. In contrast, for services, the descending order of mean values is D, B, C and A, but there is no difference between B and C. This means that a service company focuses on demand integration of customer to enhance customer satisfaction rather than efficiency. And it is not common in services since some of them do not have the supply part. And an integrated company with supplier gains more efficiency than that with customer. In reverse, an integrated company with customer acquires higher customer satisfaction than that with supplier. Therefore, this research addresses important results as follows:

• Manufacturers and services to pursue operational efficiency or customer satisfaction are able to achieve the best performance model by adopting demand chain management integration among the integration strategies.

• Supply integration provides better performance for manufacturers than demand integration in pursuing operational efficiency or customer satisfaction.

• Supply integration provides better performance for services than demand integration in pursuing operational efficiency.

• Demand integration provides better performance for services than supply integration in pursuing customer satisfaction.

• A low integration delivers the fewest benefits to manufacturers and services.

REFERENCES

- [1] Bhatnagar, R., and S. Viswanathan, "Re-engineering global supply chains", *International Journal of Physical Distribution & Logistics Management*, Vol.30, No.1, pp.13-34, 2000.
- [2] Bowersox, D.J., and P.J. Daugherty, "Logistics paradigms: the impact of information technology", *Journal of Business Logistics*, Vol.16, No.1, pp.65-80, 1995.
- [3] Cachon, G.P., and M. Fisher, "Supply Chain Inventory Management and the Value of Shared Information", *Management Science*, Vol.46, No.8, pp.1032-1048, 2001.
- [4] Carr, A.S., and J.N. Pearson, "Strategically managed buyer-supplier relationships and performance outcomes", *Journal of Operations Management*, Vol.17, No.5, pp.497-519, 1999.
- [5] Chien, T.K., T.H. Chang and C.T. Su, "Did your really win customers' satisfaction?", *Industrial Management & Data Systems*, Vol.103, No.4, pp.253-262, 2003.
- [6] Corbett, C.J., Blackburn, J.D., and Wassenbove L.N., "Case Study: Partnerships to Improve Supply Chain", *Sloan Management Review*, Vol.40, No4, pp.71-82, 1999.
- [7] Fisher, M.L., J.H. Hammond, W.R. Obermeyer, and A. Raman, "Making supply meet demand in an uncertain world", *Harvard Business Review*, Vol.72, No.3, pp.83-93, 1994.
- [8] Freeland, J.R., "A survey of just-in-time purchasing practices in the United States", *Production and Inventory Management Journal*, Vol.32, No.2, pp.43-49, 1991.
- [9] Frohlich, M.T. and R. Westbrook, "Arcs of integration: an international study of supply chain strategy", *Journal of Operations Management*, Vol.19, No.2, pp.185-200, 2001.
- [10] Frohlich, M.T. and R. Westbrook, "Demand chain management in manufacturing and service: web-based integration, driver and performance", *Journal of Operations Management*, Vol.20, No.6, pp.729-745, 2002.
- [11] Handfield, R.B., "A resource dependence perspective of just-in-time purchasing", *Journal of Operations Management*, Vol.11, pp.289-311, 1993.
- [12] Heikkila, J., "From supply to demand chain management: efficiency and customer satisfaction", *Journal of Operations Management*, Vol.20, No.6, pp.747-767, 2002.
- [13] Lancioni, R.A., M.F. Smith and T.A. Oliva, "The Role of the Internet in Supply Chain Management", *Industrial Marketing Management*, Vol. 29, pp.45-65, 2000.
- [14] Lee, H.L., V. Padmanabhan, and S. Whang, "The bullwhip effect in supply chains", *Sloan Management Review*, Vol.38, No.3, pp.93-102, 1997.
- [15] Lee, H.L. and S. Whang, "E-Business and Supply Chain Integration". *Stanford Global Supply Chain Forum*, <http://www.stanford.edu/group/scforum/>, 2001.
- [16] Marbert, V.A., and M.A. Venkataraman, "Special research focus on supply chain linkages: challenges for design and management in the 21st century", *Decision Science*, Vol.29, No.3, pp.537-552, 1998.
- [17] Narasimhan, R., and J. Jayaram, "Casual Linkages in Supply Chain Management: An Exploratory Study of North American Manufacturing Firms", *Decision Sciences*, Vol.29, No.3, pp.579-605, 1998.
- [18] Rahman, Z., "Internet-based supply chain management: using the Internet to revolutionize your business", *International Journal of Information Management*, Vol.23, pp.493-505, 2003.
- [19] Vollmann, T.E., C. Cordon, and J. Heikkila, "Teaching supply chain management to business executives", *Production and Operations Management*, Vol.9, No.1, pp.81-90, 2000.
- [20] Woods, J., "The Demand Chain", *Gartner Symposium ITXPO*, 2002.

Collaborative Planning, Forecasting and Replenishment: Demand Planning in Supply Chain Management

Hokey Min¹, Wenbin Yu²

¹UPS Center for World Wide Supply Chain Management, University of Louisville, Louisville, KY, USA

²School of Management and Information Systems, University of Missouri, Rolla, MO 65409, USA
h0min001@gwise.louisville.edu

ABSTRACT

Recognizing the value of information sharing among supply chain partners, a growing number of firms have expressed keen interest in jointly creating customer demand, managing business functions and leveraging the strength of their supply chain partners. In particular, such interest sparked the rapid development and implementation of collaborative planning, forecasting and replenishment (CPFR) that was proven to be successful in minimizing safety stocks, improving order fill rates, increasing sales, and reducing customer response time. Despite increasing popularity, key drivers for the successful development and implementation of CPFR have not been fully understood by practitioners and academicians alike. This paper unveils the invisible challenges and opportunities for adopting and implementing CPFR. Also, it provides an overview of CPFR in comparison to other alternative forecasting techniques such as agent-based forecasting and focus forecasting, while synthesizing the past evolutions and future trends of CPFR in a supply chain setting.

Key words: Collaborative Planning, Forecasting and Replenishment, supply chain management, demand planning

1. INTRODUCTION

In contrast to the traditional business paradigm that focuses on the effectiveness and efficiency of separate business functions, a growing number of firms have begun to realize the strategic benefits of jointly planning, controlling, and designing a supply chain as a whole through inter-functional and inter-organizational integration. These benefits include: reduced inventory (or working capital), less frequent rush delivery, faster product flow, quicker customer response time, higher asset utilization, larger market share, and increased revenue and profit. Despite a host of benefits gained from supply chain integration, only a few firms have adopted and successfully implemented a concept of jointly planning, controlling, and designing a supply chain. In fact, Gustin (2001) reported that less than one third (26%) of the 300 U.S. companies surveyed recognized and successfully implemented supply chain integration. Also, the respondents of the American Shipper/Mercer Management Consulting Survey conducted by Artman and Sabath (1995) indicated that supply chain integration was one of the most pressing logistics related problems. A lack of success in implementing supply chain integration is often due to poorly coordinated data exchange or information sharing among supply chain partners. Thus, the ultimate success of supply chain integration will depend on supply chain partners' willingness to share real-time information throughout the supply chain and develop joint planning processes for adding value to end-customers' needs (see, e.g., Raghunathan, 1999; Cachon and Fisher, 2000; Lee et al., 2000; Thonemann, 2002; Zhao et al., 2002 a, b for various benefits of sharing information in supply chains). An information sharing process that has gained increased popularity

over the years is collaborative planning, forecasting, and replenishment (CPFR).

In general, CPFR is referred to as a nine-step joint demand planning process that aims to enhance supply chain visibility by improving order forecasts and fulfillment through continuous communications among multiple supply chain partners. The nine-step process is comprised of: (1) develop front-end agreement; (2) create joint business plan; (3) create sales forecasts; (4) identify exceptions for sales forecasts; (5) resolve/collaborate on exception items; (6) create order forecasts; (7) identify exceptions for order forecasts; (8) resolve/collaborate on exception items; (9) order generation (see, e.g., Ackerman, 2000; Logility, 2000). The detailed steps of CPFR are graphically displayed in Figure 1. Although CPFR evolved from traditional collaborative tools, such as: electronic data interchange (EDI), vendor managed inventory (VMI), and efficient consumer response (ECR), it differs from others in that it brings mutual benefits to all the supply chain partners involved by utilizing more interactive, broader communication processes throughout the supply chain rather than relying on limited transaction-level automation. Other benefits of CPFR include: higher inventory turnover, lower stock-out rate, improved order fill rate, improved cash flow; more accurate production scheduling, more amicable business relationships among supply chain partners, reduced cycle time, reduced order picking/receiving costs, reduced labor costs, and quicker response to customer needs (Sherman, 1998; Williams 1999; Barratt and Oliveira, 2001; Langabeer and Stoughton, 2001; McKaige, 2001; McCarthy and Golicic, 2002; Andraski and Haedicke, 2003). For example, the pilot implementation of CPFR by Wal-Mart and its supplier, Warner-Lambert,

increased average in-stock rates from 87 to 98%, reduced lead time from 21 to 11 days, and increased sales volume by \$8.5 million in 1995 (Fahrenwald et al, 2001). A German office supplies manufacturer named Herlitz AG, which shared information with its retailer named Metro through CPFR, reduced its inventory by 15%, curtailed its stock-outs by 50%, and increased its annual sales by 3% (Andraski, 2000). Similarly, Johnson & Johnson reported to increase its in-stock rate from 91.5% to 93.8% after adopting CPFR (Inventory Management Report, 2002). After CPFR was implemented for Sears and Michelin in 2001, they reported that in-stock levels at the Sears stores were improved by 4.3% and Sears distribution centers-to-store fill rate was increased by 10.7%, while the combined Sears and Michelin inventory levels were reduced by 25% (Steermann, 2003). The detailed illustration of CRPF implementation success for other companies can be found in Seifert (2003).

Despite the aforementioned promises, CPFR is not without its obstacles. These obstacles may include: cultural and technical incompatibility among supply chain partners, lack of trust among supply chain partners, lack of scalability, lack of internal alignment, inadequate software and technology support, substantial start-up investment for building a communication infrastructure, antitrust laws, legacy systems, and difficulty in real-time coordination of information exchange (e.g., Mentzer et al., 2000; Barratt and Oliveira, 2001; McCarthy and Golicic, 2002; Seifert, 2003). One way of overcoming such obstacles is the integration of CPFR with other alternative forecasting tools, such as: agent-based forecasting and focus forecasting, that can complement CPFR. With this in mind, the primary purposes of this article are to:

1. Compare and contrast CPFR with other related demand forecasting and planning techniques such as agent-based forecasting and focus forecasting;
2. Synthesize the existing CPFR literature with respect to their practical implications and technical merits;
3. Develop a hierarchical taxonomy for the existing CPFR literature and categorize it according to its application area, problem scope, and methodology;
4. Summarize research trends and identify untapped research topics associated with CPFR;
5. Discuss the future outlook for extensions of existing CPFR literature.

2. CPFR VS. AGENT-BASED FORECASTING

Since a supply chain involves the synchronization of a series of inter-related but different stages of business processes influencing multiple trading partners, its demand planning and forecasting cannot rely on a single, stand-alone forecasting tool. One of the emerging concepts that fits into the supply chain framework is agent-based forecasting which was designed to

automate the coordinated planning and communication processes throughout the entire supply chain. In general, agent-based forecasting is a computer-based forecasting system that aims to facilitate enterprise-wide integration through non-proprietary message transfers and automated reasoning capabilities available from intelligent software agents. According to Reis (1999), an agent refers to an autonomous entity that can take certain actions to accomplish a set of goals and can compete and cooperate with other agents while pursuing its individual goals. An agent is often characterized by its ability to exploit significant amounts of domain knowledge, overcome erroneous input, use symbols and abstractions, learn from the environment, operate in real time, and communicate with others in natural language (Newell, 1989).

Exploiting such characteristics, an agent concept was recently introduced to handle various logistics and supply chain issues including a shop floor control within a material requirement planning (MRP), traffic control, joint production planning, and business-to-business communication and negotiation (Van Dyke Parunak, 1988; Satapathy et al., 1998; Swaminathan, 1998; Garcia-Flores et al., 2000; Fox et al., 2000; Ito and Saleh, 2000). More recently, Yu et al. (2002) developed an agent-based forecasting system to predict end-customer demand through information exchange among supply chain partners. In forecasting applications, they utilized four multiple agents with different roles: task agents; coordination agents; data collection agents; interface agents (see Figure 2 for a graphical display of these agents). Their experiments with actual data showed that agent-based forecasting performed better to predict end-customer demand than a collection of traditional time series forecasting techniques in a supply chain setting. Thus, they conclude that agent-based forecasting was suitable for demand planning in a supply chain setting as a substitute for CPFR. Though CPFR and agent-based forecasting have similarities in that both are intended to coordinate multiple sources of data for joint demand planning and forecasting, they differ in many respects. The differences between CPFR and agent-based forecasting are summarized in Table 1.

3. CPFR VS. FOCUSED FORECASTING

Focus forecasting is an expert system that identifies a simple forecasting rule-of-thumb, which worked best in the past, and uses it to make a short-term prediction for future events such as sales or customer demands. Focus forecasting is often comprised of three steps: (1) simulate the past forecasts using a variety of simple forecasting rules; (2) evaluate the performances of these simple rules with respect to forecasting errors such as mean absolute deviation (MAD) and mean absolute percentage error (MAPE); (3) select the forecasting rule that performed best as the forecasting method that will be used to forecast the next period's demand. Also, focus

forecasting is characterized by its simplicity, lack of file maintenance, and user friendliness (Smith, 1991; Smith, 1997). Focus forecasting is different from CPFR in that the former does not involve trading partners for making forecasts and focuses on internal forecasts within the company, whereas the latter shares information with the multiple supply chain partners and develop both intra- and inter-organizational demand plans (see Table 1 for detailed distinctions between CPFR and focus forecasting).

4. TAXONOMY OF THE CPFR LITERATURE

The CPFR taxonomy to be developed here uses three broad classification schemes: (1) *problem scope* as a criterion for measuring the breadth and depth of the CPFR study, (2) the *methodology* as a criterion for evaluating the theoretical advance of the CPFR study, and (3) the *implementation status* for assessing the practicality of the CPFR study.

4.1 Problem scope

The problem scope is categorized with respect to the level of CPFR integration: (1) *intra-organizational integration* across different business functions; (2) *inter-organizational integration* across vertical supply chain links. The inter-organizational integration can be further sub-classified into: (1) *two-tier* integration; (2) *n-tier* integration. A two-tier integration focuses on the collaborative relationship between two primary supply chain partners (typically the manufacturer and retailer); whereas, a n-tier integration involves more than two (multiple) supply chain partners across the industry and vertically extends information sharing mechanism to minimize “bullwhip” effects as illustrated in Figure 3. The bullwhip effect is generally referred to as an inverse ripple effect of forecast errors throughout the supply chain that often leads to amplified supply and demand misalignment where orders to the upstream supply chain partner tend to exaggerate the true patterns of end-customer demand since each supply chain partner’s view of true demand can be blocked by its immediate downstream supply chain partners (see, e.g., Lee et al. 1997; Min, 2000). Either two-tier or n-tier integration can also be subdivided into one-to-one, one-to-many, and many-to-many integration (e.g., CPFR partnerships among multiple manufacturers and multiple retailers) depending on the number of horizontal supply chain partners involved in CPFR.

4.2 Methodology

The broad subcategories of methodological classification are descriptive (conceptual) and normative (analytical) studies. The descriptive studies often illustrate the numerous managerial benefits of using CPFR or evaluate the outcomes of CPFR in comparison to traditional stand-alone demand planning through case examples. On the other hand, normative studies

designed quantitative models to assess the positive impact of CPFR (or information sharing across the vertical supply chain) on various supply chain performances in comparison to old legacy systems or less structured forecasting procedures. These quantitative models can be broken down into mathematical models and simulation experiments. The core mathematical models also include various forecasting techniques which may be categorized as: time series and causal methods. Time series (univariate forecasting) methods assume that the past is an adequate representation of the future and that a product’s past demand pattern can be extrapolated to predict future demand (Tyagi, 2002). Causal (multivariate forecasting) methods make predictions of the future by modeling the relationship between a series and other series (DeLurgio, 1998). Time series methods include moving average, exponential smoothing, decomposition, linear trends, and Box-Jenkins, while causal methods include regression and econometric models (Chambers et al., 1971; DeLurgio, 1998).

4.3 Implementation status

Since supply chain managers may be interested in determining the applicability of the proposed model and procedure, we included the third dimension of the taxonomy indicating whether or not the proposed model and procedure was implemented, and whether or not the model was developed based on actual data.

5. SYNTHESIS AND FUTURE RESEARCH DIRECTIONS

Despite the relatively short history of CPFR, we have noticed a clear research stream established by many pioneering scholars who initiated CPFR studies during the last few years. A few of these studies worth noting include: (1) Aviv (2001) who developed a mathematical model as a means of measuring the magnitude of the benefit of joint demand forecasting and inventory planning on supply chain performance (e.g., supply chain cost); (2) Barratt and Oliveira (2001) who identified key inhibitors for CPFR implementation through an empirical survey; (3) McCarthy and Golobic (2002) who proposed managerial guidelines for successful implementation of CPFR based on their case studies; (4) Chen et al. (2000) who introduced a mathematical framework that helped measure the impact of demand forecasting on order variability (or bullwhip effect) in supply chains. Regardless of methodological differences among these studies, all of these studies share a common theme -- a verification of the positive impact of CPFR on supply chain performance (see Table 2 for detailed summaries of past CPFR studies). Following this theme, Zhao et al. (2001), Zhao and Xie (2002), Zhao et al. (2002a, 2002b), and Dejonckheere et al. (2002) attempted to evaluate the impact of demand forecasting on supply chain performance including the bullwhip phenomena using simulation models. Most of

these models, however, are confined to two-tier supply chain integration and focused their attention on supply chain costs in measuring the supply chain performance. With many open research questions unanswered and buoyed by increasing interest in CPFR, additional research is needed to consider the following areas:

- As visibility increases proportionately to the number of tiers participating in inter-organizational collaboration, future research efforts should be directed toward the development of n-tier CPFR that can support multiple supply chain partners and multiple industries. In other words, the scope of CPFR should be extended to include distributors (e.g., carriers and 3PLs) and multi-tier suppliers, while encompassing the elements of collaborative product design, transportation, event and categorical management. In particular, the integration of collaborative transportation management (CTM) into CPFR or vice versa may enhance logistics efficiency by increasing freight consolidation opportunities, updating shipment status on a real-time basis, and utilizing trucking capacity. Thus, the integration of CTM with CPFR can be an intriguing subject of future research.
- Although the potential benefits of CPFR were listed in the past literature, its success in a real-world setting has not been well documented. Additional case studies reporting the CPFR success and failure stories are needed to convince CPFR critics to consider CPFR as a major demand forecasting and planning tool. In particular, these studies should identify key enablers and impediments for the successful implementation of CPFR.
- In evaluating the magnitude of CPFR benefits, most of the existing literature relied on traditional supply chain performance measures such as inventory carrying costs and backorder costs. The development of more specific supply chain performance metrics for measuring CPFR benefits should be in order for future research. These metrics may include: cash-to-cash cycle time, inventory turns, order fill rates, and stock-out rates (see, e.g., Min and Zhou 2002 for a detailed discussion of supply chain performance metrics).
- Since CPFR is intended to bring joint benefits for all the supply chain partners involved in the demand forecasting and planning process, future research endeavors may focus on the verification of joint benefits gained from CPFR during the multiple time periods. That is to say, more longitudinal studies should be conducted to examine the long-term effects of CPFR on supply chain connectivity and visibility.
- Given the evolution of agent-based forecasting and focus forecasting that can replace or complement CPFR, the comparative analyses of CPFR versus agent-based forecasting or focus forecasting may be a fruitful area for future research. Similarly, the complementary nature of CPFR and ECR or CPFR and VMI needs to be investigated to gain further insights into the role of ECR or VMI in CPFR.

ACKNOWLEDGMENT

The authors would like to express sincere gratitude to the UPS Foundation for partly funding this research.

REFERENCES

1. Ackerman, K.B. (2000), "CPFR-How it could change the warehouse," *Warehousing Forum* 15(10), 1-2.
2. Andraski, J.C. (2000), "Logistics Strategies for the New Millennium, CPFR, the Here and Now," presented at the *Warehousing Education and Research Council Conference*, San Antonio, Texas, May 3.
3. Andraski, J.C. and Haedicke, J. (2003), "CPFR: Time for the Breakthrough?" *Supply Chain Management Review* 7(3), 54-60.
4. Aviv, Y. (2001), "The effect of collaborative forecasting on supply chain performance," *Management Science* 47(10), 1326-1343.
5. Artman, L.B. and Sabath, R.E. (1995), "Are you ready for change?" *American Shipper* 37(2), 42-48.
6. Barratt, M., and Oliveira, A. (2001), "Exploring the experiences of collaborative planning initiatives," *International Journal of Physical Distribution & Logistics Management* 31(4), 266-289.
7. Cachon, G.P. and Fisher, M. (2000), "Supply chain inventory management and the value of shared information," *Management Science* 46, 1032-1048.
8. Cachon, G.P. and Lariviere, M.A. (2001), "Contracting to assure supply: How to share demand forecasts in a supply chain," *Management Sciences* 47(5), 629-646.
9. Chambers, J.C., Mullick, S.K., and Smith, D.D. (1971), "How to choose the right forecasting techniques," *Harvard Business Review*, 49(4), 45-74.
10. Chen, F., Drezner, Z., Ryan, J. K. and Simchi-Levi, D. (2000), "Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information," *Management Science* 46(3), 436-43.
11. Dejonckheere, J., Disney, S. M., Lambrecht, M. R., and Towill, D.R. (2002), "Transfer function analysis of forecasting induced bullwhip in supply chains," *International Journal of Production Economics* 78(2), 133-144.
12. DeLurgio, S.A. (1998), *Forecasting Principles and Applications*, New York, NY: Irwin-McGraw-Hill.
13. Fahrenwald, B., Wise, D., and Glynn, D. (2001), "Supply chain collaboration: Close encounters of the best kind," *Business Week (special advertising section)*, March 21, 1-14.
14. Foote, P.S. and Krishnamurthi, M. (2001), "Forecasting using data warehousing model: Wal-Mart's experience," *The Journal of Business Forecasting* 19, 13-17.
15. Fox, M.S., Barbiceanu, M., and Teigen, R. (2000), "Agent-oriented supply chain management," *International Journal of Flexible Manufacturing Systems* 12, 165-188.
16. Garcia-Flores, R., Wang, X.Z., and Goltz, G.E. (2000), "Agent-based information flow for process industries' supply chain modeling," *Computers and Chemical Engineering* 24, 1135-1141.
17. Gustin, C.M. (2001), "Supply chain integration: Reality or myth?" *Distribution Business Management Journal* 1(1), 71-74.

18. Inventory Management Report (2002), "CPFR-No longer a question of if but when," *IOMA Inventory Management Report: Improving Logistics and Supply Chain Management*, December, 1-14.
19. Ireland, R. and Bruce, R. (2000), "CPFR: Only the beginning of collaboration," *Supply Chain Management Review* 4(4), 80-88.
20. Ito, T. and Saleh, M.R. (2000), "A blackboard-based negotiation for collaborative supply chain systems," *Journal of Material Processing Technology* 107, 398-403.
21. Langabeer, J. and Stoughton, T. (2001), "Demand planning and forecasting in the high technology industry," *The Journal of Business Forecasting* 19, 7-10.
22. Lee, H.L., So, K.C., and Tang, C.S. (2000), "The value of information sharing in a two-level supply chain," *Management Science* 46, 626-643.
23. Lee, H.L., Padmanabhan, V., and Whang, S. (1997), "Information distortion in a supply chain: the bullwhip effect," *Management Science* 43, 546-558.
24. Logility (2000), *Collaborative Transportation Management: An Executive White Paper*, Atlanta, GA: Logility, Inc.
25. McCarthy, T.M. and Golicic, S.L. (2002), "Implementing collaborative forecasting to improve supply chain performance," *International Journal of Physical Distribution & Logistics Management* 32(6), 431-454.
26. McKaige, W. (2001), "Collaborating on the supply chain," *IIE Solutions*, 33(3), 34-37.
27. Mentzer, J.T., Foggini, J.H., and Golicic, S.L. (2000), "Collaboration: The enablers, impediments, and benefits," *Supply Chain Management Review* 4, 52-58.
28. Min, H. (2000), "The bullwhip effect in supply chain management," in *Encyclopedia of Production and Manufacturing Management* edited by P. Swamidass, Norwell, MA: Kluwer Academic Publishers, 66-70.
29. Min, H. and Zhou, G. (2002), "Supply chain modeling: past, present and future," *Computers & Industrial Engineering* 43(1), 231-249.
30. Newell, A. (1989), "Putting it all together," in *Complex Information Processing: The Impact of Herbert A. Simon*, edited by D. Klahr and K. Kotovsky, Hillsdale, NJ: Lawrence Erlbaum Associates.
31. Raghunathan, S. (1999), "Interorganizational collaborative forecasting and replenishment systems and supply chain implications," *Decision Sciences* 30(4), 1053-1071.
32. Reis, B.Y. (1999), "A multi-agent system for on-line modeling, parsing and prediction of discrete time series data," in *Computational Intelligence for Modeling, Control & Automation* edited by M. Mohammadian, IOS Press, 164-169.
33. Satapathy, G., Kumara, S.R.T., and Moore, L.M. (1998), "Distributed intelligent architecture for logistics (DIAL)," *Expert Systems with Applications* 14, 409-424.
34. Sherman, R. J. (1998), "Collaborative planning, forecasting & replenishment (CPFR): Realizing the promise of efficient consumer response through collaborative technology," *Journal of Marketing Theory and Practice* 6(4), 6-9.
35. Seifert, D. (2003), *Collaborative Planning, Forecasting, and Replenishment: How to Create a Supply Chain Advantage*, New York, NY: AMACOM.
36. Smith, B.T. (1991), *Focus Forecasting and DRP: Logistics Tools of the Twenty-First Century*, New York, NY: Vantage Press.
37. Smith, B.T. (1997), *Focus Forecasting: Computer Techniques for Inventory Control Revised for the Twenty-First Century*, Fredericksburg, VA: Book Crafters.
38. Steermann, H. (2003), "A Practical Look at CPFR: The Sears-Michelin Experience," *Supply Chain Management Review* 7(4), 46-53.
39. Swaminathan, J.M. (1998), "Modeling supply chain dynamics: A multiagent approach," *Decision Sciences* 29(3), 607-632.
40. Thonemann, U.W. (2002), "Improving supply-chain performance by sharing advance demand information," *European Journal of Operational Research* 142(1), 81-107.
41. Tyagi, R. (2002), "How to evaluate a demand planning and forecasting package," *Supply Chain Management Review* 6(5), 48-56.
42. Van Dyke Parunak, H. (1998), *Practical and Industrial Applications of Agent-based Systems*, White Paper, Ann Arbor, MI: Industrial Technology Institute.
43. Williams, S.H. (1999), "Collaborative planning, forecasting, and replenishment," *Hospital Material Management Quarterly* 21(2), 44-51.
44. Yu, W., Graham, J.H., and Min, H. (2002), "Dynamic pattern matching for demand forecasting using temporal data mining," *Proceedings of the 2nd International Conference on Electronic Business*, December, 400-402.
45. Zhao, X. and Xie, J. (2002), "Forecasting errors and the value of information sharing in a supply chain" *International Journal of Production Research* 40(2), 311-335.
46. Zhao, X., Xie, J., and Lau, R.S.M. (2001), "Improving the supply chain performance: use of forecasting models versus early order commitments," *International Journal of Production Research* 39(17), 3923-3939.
47. Zhao, X., Xie, J., and Leung, J. (2002a), "The impact of forecasting model selection on the value of information sharing in a supply chain," *European Journal of Operational Research* 142(2), 321-344.
48. Zhao, X., Xie, J., and Wei, J.C. (2002b), "The impact of forecast errors on early order commitment in a supply chain," *Decision Sciences* 33(2), 251-280.

THE COMPLETE TABLES AND FIGURES WILL BE AVAILABLE FROM THE FIRST AUTHOR UPON REQUEST.

Cooperation in Reverse Logistics

Yunfeng Zhang, Jianjing Sun

School of Management, Xiamen University, Fujian, Xiamen 361005, China
yfzhang@126.com

ABSTRACT

Many businesses have now realized that a Reverse Logistics system can be used to gain competitive advantage. Even the best retailers and vendors now have reason to review how their returns programs are managed. New technologies and service providers now offer an opportunity to improve overall 'total returns performance'. This means not only reducing the 'total cost of returns', but utilizing returns to improve customer loyalty, merchandising and product performance. Improving 'total returns performance' can transform an increasingly costly and complex process into a competitive advantage. It has forced businesses to reengineer their business processes and look into what can be the next practices in business rather than adopt best practices. Reverse Logistics is a very complex and specialized area of any supply chain and it involves handling individual incoming parcels, opening and inspecting products, communicating with internal departments, customers and vendors and then directing products into disposition channels which will provide the highest value. It is necessary to have a synchronized effort from all the supply chain partners for Reverse Logistics to be successful. This paper explains the various elements of reverse logistics, the need for cooperation in reverse logistics and the opportunities which arise through the processes of Reverse Logistics.

Keywords: *reverse logistics, cooperation, supply chain*

1. INTRODUCTION

Reverse Logistics by its nature is a very complex and specialized area of any supply chain. No matter what the product is, how it is sold, or who the customers are, every company needs to focus on recovering the greatest value from returns. All of this while maintaining customer loyalty, cooperation cost and harvesting information that helps reduce returns in the future. Cooperation between all supply chain partners is the key to establish the smooth flow of products and information in reverse logistics. Returns require special information systems, dedicated equipment and personnel trained in non-standard processes. Returns processing involves handling individual incoming parcels, opening and inspecting products, communicating with internal departments, customers and vendors and then directing products into disposition channels, which will provide the highest value. To create visibility in the returns supply chain a firm requires greater information sharing and using automated notification to support integrated decision making.

2. WHAT IS REVERSE LOGISTICS?

Reverse logistics, that is, all operations related to the extension of useful lifecycle for used products, commercial returns, excess inventory and packaging materials, gains increasing attention globally for its

promising financial potentials, its sustainable growth alternative and the environmental positive impact it could have. Reverse Logistics is a rather general term. In the broadest sense, Reverse Logistics stands for all operations related to the reuse of products and materials. The management of these operations can be referred to as Product Recovery Management (PRM). PRM is concerned with the care for products and materials after they have been used. Some of these activities are, to some extent, similar to those occurring in case of internal returns of defective items due to unreliable production processes. Reverse Logistics though refers to all logistic activities to collect, disassemble and process used products, product parts, and/or materials in order to ensure a sustainable (environmentally friendly) recovery.

Reverse logistics deals with five basic questions:

1. What alternatives are available to recover products, product parts, and materials?
2. Who should perform the various recovery activities?
3. How should the various activities be performed?
4. Is it possible to integrate the activities that are typical for reverse logistics with classical production and distribution systems?
5. What are the costs and benefits of reverse logistics, both from an economical as an environmental point of view?

Below is an example of the reverse logistics:

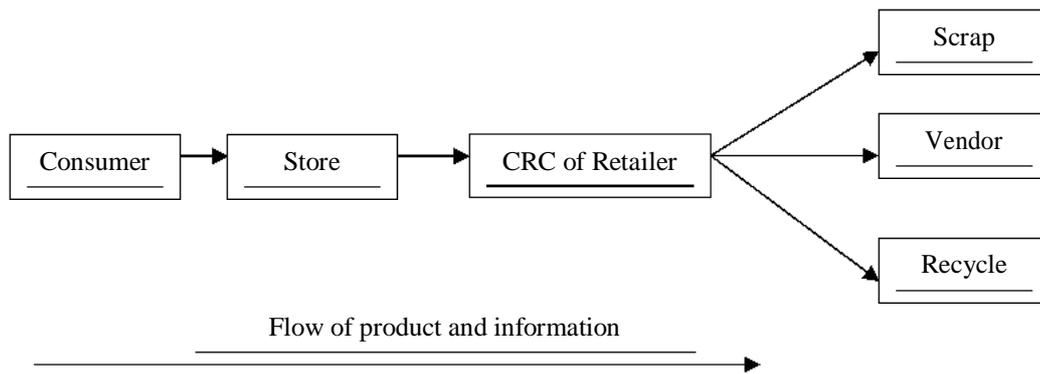


FIGURE 1: REVERSE LOGISTICS IN RETAIL INDUSTRY

3. FACTORS IN REVERSE LOGISTICS

3.1 Zero returns

Zero Returns is a program where the company in question does not accept returns from its customer. Rather, it gives the retailer an allowable return rate, and proposes guidelines as to the proper disposition of the items. This strategy basically passes the returns responsibility onto the retailer, while reducing costs for the manufacturer or distributor.

3.2 Asset recovery

Asset recovery is the “classification and disposition of returned goods as surplus, obsolete, scrap, waste and excess material products, and other assets, in a way that maximizes returns to the owner, while minimizing costs and liabilities associated with the dispositions.”

3.3 Outsourcing

Many companies have established sound reverse logistics systems and procedures. In many cases, however, it makes more sense for the firm to outsource their reverse logistics functions than keep those in-house.

3.4 Gate keeping

Gate keeping has been defined as “the screening of defective and unwarranted returned merchandise at the entry point into the reverse logistics process.”

3.6 Compacting disposition cycle time

Disposition cycle time is the amount of time needed to figure out what to do with returned products once they arrive. Often, when material often comes back in to a distribution center, it is not clear whether the items are defective, can be reused or refurbished, or need to be sent to a landfill.

3.7 Reverse logistics information systems

With the emphasis being given on Reverse Logistics today, the IT system developed to handle reverse logistics should be flexible enough for inevitable future expansion, as well as to have the ability to handle the many exceptions involved in reverse logistics. The reverse logistics system should create a database and track returns at the store level back through the pipeline to the manufacturer.

3.8 Centralized Return Centers

CRC’s are dedicated specialized buildings to handle returns. All of the company’s customers send their returns to such a central location where the items are subsequently sorted, processed, and “disposed.” This system is essential for orderly and organized method of handling returns.

4. WHAT IS SUPPLY CHAIN COOPERATION?

Cooperation = Getting to Know You, Much Better

All trading partners have a transactional relationship, and some even share lots of information about their businesses. Cooperation goes much further to take the uncertainty out of future product supply/demand through joint planning and action. By following a specified methodology – known as Cooperation Planning, Forecasting and Replenishment (CPFR) – companies can dramatically improve supply chain effectiveness with new product and package design, demand planning, synchronized production scheduling, and logistics planning. According to the CGE&Y survey mentioned earlier, this practice is catching on: 25% of respondents indicated that CPFR has been implemented in their firms. CPFR’s popularity is spreading beyond company-to-company initiatives to include industry consortia as well. Supply chain visibility and cooperation capabilities are evolving rapidly, and, not surprisingly, logistics experts often disagree on where this whole effort is headed.

5. WHY COOPERATION IN REVERSE LOGISTICS?

5.1 Return-to-vendor adds significant and unnecessary cost.

Retailers in many industries simply send returned products back to vendor/distributors for credit. Devaluation of product as it moves backward slowly in the supply chain is dramatic. Handling, storage and shrinkage costs mount at every step of the way as the product becomes increasingly out of date. The administrative costs of managing the complex, diverse returns policies among hundreds, or even thousands of retailers, distributors and manufacturers is another major cost of return-to-vendor programs.

In an effort to reclaim some of these costs, leading retailers are now negotiating attractive “no returns” pricing with their suppliers in exchange for taking over all returns, including warranty and repairs.

5.2 Merchants going online are finding returns a challenge.

With this boom in online retailing has come some chaos in all areas of customer service and fulfillment, particularly with returns. According to data from a number of industry research firms, return rates for online sales are substantially higher than traditional retailing. Rates of 20% to beyond 30% for certain categories have been reported. Most fulfillment centers, information systems and customer service centers designed for handling traditional retail simply cannot efficiently process this volume of returned merchandise. As a result, returns costs are high and returning credit to consumers has often been slow. This poor performance is a barrier to new or repeat sales to time-focused buyers.

5.3 The total cost of returns is high.

The total costs associated with returns are significant. These costs are cross-functional and hit numerous expense lines, cost centers and budgets. As a result, retailers and manufacturers have difficulty measuring all of the hidden costs or fully understanding the business impact. The hard costs can, with effort, be measured. These include warehouse staff, customer service associates, shipping, storage and inventory space, third party vendors, packaging for disposition, disposal fees and other direct costs. Higher prices from suppliers for accepting returns can also be identified. The soft costs, such as customer satisfaction and loyalty, are much more difficult to measure but can have a major impact nonetheless. Retailers interviewed recently believe that the total cost of returns could be as much as 5-10% of sales.

5.4 Value recovery from returns has traditionally been low.

There are a number of established methods for recapturing some value from returned merchandise. In-store sales and return-to stock are labor intensive, provide mixed results and also pose risks to brand

equity and cannibalized sales. Traditional liquidation channels have provided low value recovery—often only 10%-20% of retail value due to slow and inefficient paths back to consumers. The Internet has enabled new channels for selected merchandise that offer more direct, faster resale paths with higher recovery.

6. Why focus on Reverse Logistics?

Following are a few forceful reasons to focus on Reverse logistics:

6.1 Return rates are higher than ever.

Return rates are high and climbing, especially for online businesses which are experiencing increased return rates as a result of trial and impulse Internet purchasing. Calls and traffic to customer service also increases as online buyers, in particular, expect fast credits and refunds.

6.2 The cost of returning items to vendors is ultimately absorbed by retailers.

Retailers in many industries simply return products to vendors. Sending more products back to vendor's results in necessary price increases to their customers, the distributors and retailers.

6.3 Do third party logistic providers handle returns efficiently?

Some retailers have successfully outsourced fulfillment and logistics functions to third parties for better service, lower costs and increased flexibility. However, few of these providers have yet to develop the special processes or to make the incremental investments necessary to handle returns efficiently.

6.4 Taking full-Advantage of new returns technology is the key.

New technologies and service providers have emerged to address the increasing demand for better solutions in returns management.

6.5 Returns as a competitive advantage.

Retailing has become competitive, so every opportunity should be taken for bottom-line improvements. Returns have the potential to be one of the greatest untapped sources for increased revenue and customer loyalty.

6.6 Returns: not just a necessary cost of doing business.

Returns have traditionally been considered just a necessary cost of business. However, the costs associated with returns are far from trivial when all of the hard and soft costs are calculated to determine true business impact.

6.7 The traditional liquidation of returns channels are slow, inefficient and costly.

Recovering value from returns is a difficult task through the traditional liquidation channels. Although these channels are widely used, they lack ease and efficiency, and the end result for merchants is the recovery of a mere 10 to 20 percent of cost. Products lose value at every step to final disposition.

6.8 Returns contain a wealth of information on products and merchandising.

Returns offer an immense amount of information about consumers and products that few retailers and merchants capture. For example, returns can provide information about the original merchandising, actual product performance, ease of use, product defects and consumer expectations.

7. COOPERATION IN REVERSE LOGISTICS

7.1 Warranty

The first steps towards collaboration need to be taken while dealing with warranty at the *customer-retailer interface*. Any product with a lapsed warranty should not be allowed into the reverse supply chain by retailers: such products consume time and effort in the reverse supply chain. If the product enters the supply chain it needs to be dispositioned – recycled, refurbished or returned deciding the appropriate course of action is a potential area of conflict at the retailer-vendor interface. Verification of products at the store level may be done using invoices or serial number of product. However, this is easier said than done because tracking of millions of individual items requires maintaining a large database which proves inefficient, retrieval of information is time consuming and not cost effective.

The next stage for collaboration comes at the *retailer-vendor interface*. Once the product is accepted by the store it is moved to the CRC of the Retailer. There generally is some time gap between these two events. This delay may be due to a minimum return quantity stipulated by the retailer to stores or specified time periods during which returns are accepted. For any product accepted close to the end of warranty, its warranty may expire during this time delay. Hence, information systems at the CRC should be responsive and raise exceptions and alerts to warn the retailer and expedite the returns process. The system should check if demand for this product exists elsewhere in the network of stores and accordingly arrange for movement of stocks. FIGURE 2 show the details.

7.2 Return Material Authorization (RMA)

Cooperation would be required between vendors and

retailers to ensure the flow of products in reverse logistics. It is important that they understand specifics of the return supply chain and agree on disposition of returned products. To ensure this cooperation information on daily returns needs to be available centrally to both the retailer and vendor so that they may optimize the frequency and cost of transportation.

7.3 Return price rationalization

For any product in the retail chain two prices come into play. First, the price paid by the retailer to the vendor and second, the price a customer pays to buy the product.

Customer-Retailer interface: On returning a product, customers demand a reimbursement from the retailer. The retailer is hence obliged to pay the customer at the price at which the product was bought. However, customers may have misplaced or lost the invoice making it impossible to verify the price for reimbursement. Two cases emerge from this scenario:

(i). Reduction in price of product after it was first purchased. Owing to lack of information, customers are reimbursed at the reduced price, they suffer a loss. This leads to loss of face of the retailer and a loss in customer loyalty and brand image of the retailer.

(ii). Reduction in price of product after it was first purchased. In this case the retailer is at a loss since he has to reimburse the customer at a higher price. The retailer hence suffers a financial loss on the return in question.

Hence, it is important to be able to track the price of individual products while being returned by customers. Serial number tracking could be the answer to resolve this issue.

Retailer-vendor interface: Once the product is in the reverse supply chain and the retailer needs to return the product to the vendor an important question arises, at what price should the vendor reimburse the retailer?

7.4 Product returns with damaged packaging

Many a time products returned by customers to retailers or vendors are not in the original package. Either the package in which the product is returned is damaged or is of a different product all together. This creates problems in reverse logistics since any identification put on the product is lost and products are not in a position to be identified by any scanning media leading to a discontinuity in information flow in the supply chain.

At the retailer-customer interface: Retailers may refuse to accept products not returned in the original packaging, this leads to friction between customers and retailers which ultimately results in an irate customer thereby affecting customer satisfaction, retention and loyalty and diminishing the brand equity of the retailer.

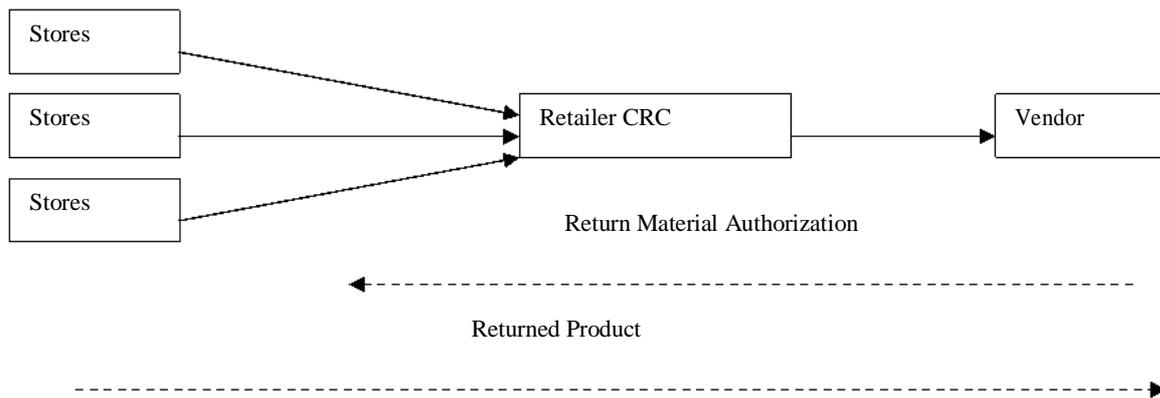


FIGURE 2: FLOWS IN SUPPLY CHAIN

8. CONCLUSION

Reverse Logistics has been stretching out worldwide, involving all the layers of supply chains in various industry sectors. While some actors in the chain have been forced to take products back, others have pro-actively done so, attracted by the value in used products. One way or the other, therefore, Cooperation in Reverse Logistics has become a key competence in modern supply chains.

ACKNOWLEDGEMENT

ICEB2004 is supported by the National Natural Science Foundation of China.

REFERENCES

[1] A. I. Kokkinaki, R. Dekker, J. van-Nunen and C. Pappis, "An Exploratory Study on Electronic

Commerce for Reverse Logistics", Supply Chain Forum, An International Journal, vol. 1, no. 1, pp: 10-17, 2000.

[2] Carter, C. R. and L. M. Ellram "Reverse logistics: A review of the literature and framework for future investigation" , Journal of Business Logistics, Vol.19 No.1, 1998.

[3] Norbridge(2004) "Supply Chain Collaboration and Visibility", www.businessweek.com

[4] Tibben-Lembke. R. S. (1998) Going Backwards: Reverse Logistics Trends and Practices. Reverse Logistics Executive Council, www.rlec.org

[5] Wipro Technologies Inc, www.wipro.com

[6] Fleischmann, M., M. Jacqueline, M. Bloemhof-Ruwaard, R. Dekker, E. Vander Laan, J. Van Nunen and L. Van wassenhove "Quantitative models for reverse logistics: A review", European Journal of Operation Research Vol.103 No.1, 1997

Coordination Mechanisms in the Supply Chain Based E-business

Li Sui-cheng^{1,2}, Cui Xian-xia¹, Liang Gong-qian²

¹Management School, Xi'an University of Technology, Xi'an 710048, China

²Management School, Northwestern Polytechnical University, Xi'an 710072, China
lisc@xaut.edu.cn, cuixianxia@163.com

ABSTRACT

Supply chain management has been influenced greatly by e-business, so does coordination which is regarded as the core of supply chain management. The characteristics of supply chain management based on e-business are analyzed firstly, which describe that information transmission has been accelerated and the relation between enterprises has become not only competitive but also cooperative under e-business environment. Differences of coordination mechanism between tradition environment and e-business environment are compared secondly. Finally, according to market relation and competitive-cooperative relation between enterprises, the suitable coordination mechanisms based on e-business are put forward. The paper presents a framework to assist managers in adopting the suitable coordination mechanisms under different business environment.

Keywords: Supply chain management, E-business, Coordination mechanism, Cooperative relation

1. INTRODUCTION

Supply chain management (SCM) has gained significance as one of the 21st century manufacturing paradigms for improving organizational competitiveness. Emphasizing strategic collaboration, information sharing, responding to market rapidly etc, the key of supply chain management is to manage and coordinate the cooperation between enterprises of supply chain. However, because the application of the network and information technology lags behind the relative advanced management mode, the traditional supply chain based on paper, fax, is difficult to share information simultaneously between enterprises. Though some manufacturing companies have adopted the systems, such as MRPII, ERP, CRM, DSS, etc, the systems can't fully support and reflect the strategic advantage and systematic characteristic of supply chain management. These systems are often limited in single enterprise and the Online Electronic Linkage between partners and the interface between the enterprise and customer are weak, which lead to the lone island of information of supply chain.

E-business is a kind of social form of business operation using modern computer information technology, especially network technology. The basic purpose of e-business is to realize the maximization of the social wealth through improving productivity, reducing the operating costs, optimizing resource distribution. The rise of e-business is a revolutionary of the business operation mode caused by the rapid development of information technology, which leads to the great change in existence foundation, operation way, management mechanism of traditional economic activity, then the traditional market ideas, and then the theories and methods of supply chain management. As a result, e-supply chain management appears being the combination of e-business and supply chain

management.

As one of new management modes, supply chain management integrates and coordinates resources of all nodal enterprises, so coordination is the core of supply chain management. Under e-business environment, the communication in organization becomes more convenient. And the institutional structure is a hierarchical network system, which make the range of management bigger and the amount of information that the members need to accept and deal with older. In addition, carrying on customer relation management and establishing learning organization make it difficult to obtain, share and integrate the information in the supply chain. Because the uncertainty of market is increasing under e-business environment and the relation between partners becomes complicated, partners will often intend to conceal the information or the knowledge resources in consideration of one's own interests. At the same time, with the application of e-business, managers can set up various coordination mechanisms through network. Using the appropriate coordination mechanisms offers the potential to help managers improve business performance, increase revenues, reduce costs and reach new markets. However, for these to happen, management practices must be adapted to the new environment. The Internet facilitates the implementation of a variety of coordination mechanisms, such as auctions, electronic agents, etc. The challenge that the managers face is to choose the mechanism that best fits each situation. So grasping the coordination mechanisms of supply chain based on e-business is very important to go on the deeper research of supply chain management and has great directive significance to apply the thoughts of supply chain management to our country.

The purpose of this paper is to identify which coordination mechanism is best fit to the practical business environment. At first the characteristics of

supply chain under e-business environment are analyzed. The differences of coordination mechanisms between tradition environment and e-business environment are compared from running environment, feedback mechanism etc. Then the suitable coordination mechanisms under e-business environment are put forward and compared. The text is summarized finally.

2 SUPPLY CHAIN MANAGEMENT BASED ON E-BUSINESS

With the improvement of information technology and the aggravation of economic globalization, the development of e-business is very fast. As the statistics of relevant materials showing, the global turnover of e-business reaches 15,300 million Yuan in 1998, rises to 26,500 million Yuan in 1999, increases about 73% per year on average. E-business is concerned by all walks of life, and management theories and operation modes have changed deeply.

2.1 Characteristics of Supply Chain Management Based on E-business

Supply chain management has been influenced by e-business. Manufacturers can sell products to consumers directly not having dealer or the retailer through e-business, which eliminate some unnecessary links of supply chain. Meanwhile, the information flow and fund flow between enterprises are more frequent. The relation between enterprises has changed. Therefore, under e-business environment, supply chain management has some new characteristics.

1) The complication of relation between enterprises

Traditional supply chain regards production as the centre. Because of the existence of the seller's market, the seller can control the buyer through price and quantity of supply, and the relation between buyer and seller is a kind of short-term contract, that is a kind of competitive relation. Because of the appearance of advanced management thought, such as MRP, JIT, TQM, etc., the relation between enterprises changes into the logistic relation progressively. The main characteristic of logistic stage is the integration of enterprises based on material. The cooperative relation of enterprises based on logistics is only a basic cooperation.

With the appearance of e-business, the cooperation between enterprises of supply chain has been strengthened. Not only the material flow, but also the information flow is transferred between enterprises. Strategic cooperative relation appears. The relation between enterprises of supply chain is becoming more complicated which is not only competitive but also cooperative.

2) Enhancement of information sharing

The purpose of information integration is to realize information sharing of supply chain enterprises, reduce the influences of uncertain factors. The information transmission of traditional supply chain is a chain which can easily lead to "bull whip effect". After e-business appears, supply chain management is set up a centre of information management based on Internet which is a mode of network information integration. It can store all kinds of data, and deal with information in time. It is helpful to realize the cooperation between enterprises and improve the integration of information in the supply chain.

3) Realization of synchronized operation

The synchronized operation between enterprises of supply chain can reduce the delay of demand information, improve the working efficiency. Because of the distorting of information transmission and the relation of competition between enterprises, the traditional supply chain is very difficult to take operation in step. Under the environment of e-business, the synchronization can be realized by Internet. Through Internet, enterprises can forecast, make and implement the relevant plan together. Meanwhile, buyers and sellers can share the result of forecasting, exchange the views and coordinate the difference of both sides in time. Therefore, supply chain management can realize the synchronization entirely on internet.

4) The transformation of management focus

Under the traditional mode of management, the focus of supply chain management is how to create value using the advantage of production technology and management. In e-business era, the focus of supply chain management is how to assure "power origin" status of consumers, how to obtain knowledge through an analysis of uncertain demand, and how to create value by sharing knowledge with the whole supply chain. So the supply chain enterprises should collect and share of demand information, and change them into knowledge which can improve the efficiency of supply chain by the knowledge converter.

5) pull-operational mode of supply chain

The operational mode of traditional supply chain is a typical kind of push management. After having produced goods, manufacturers send them to retailer or customer by utilizing logistics. In this course, the trade flow and material flow are push management. The logistics doesn't create value by itself but only plays the function of supporting while the whole goods move. However, supply chain based on e-business is different. Because the activities of commodity production, distribution, storage and delivery are all operated according to the customer's order, all the activities including trade flow, material flow and fund flow are all connected with market. Logistics has not only offered a

strong guarantee for trade flow, but also created value at the same time because it is the component of the customer service.

6) The appearance of virtual market

In e-business era, the living space of enterprises has changed to not only physical market but also virtual market. The virtual market is based on information, and controlled by information. The appearance of virtual market has changed the situation that consumers must use or enjoy products / serve through the physical market, and weakened the restriction that producers must gain and produce resources through the physical market. Physical value chain which enterprises depend on in the supply chain becomes virtual. Enterprises can seek new points of creatable value by the information technology such as integration, organization, choice and synthesizing etc.

2.2 Comparison of Coordination Mechanisms between Traditional Supply Chain and E-business Supply Chain

Coordination is the administrative behavior based on the communication of information. With the appearance of e-business, the communication of supply chain enterprises becomes more convenient. A lot of new information sharing platforms have been used. Many new coordination mechanisms are presented, such as on-line auction, electronic agent and VMI etc. At the same time, because of the improvement of information technology and the expansion of economic globalization, the relation between supply chain enterprises has not only competitive but also cooperative. So there are differences of coordination mechanisms between traditional supply chain and e-business supply chain.

1) Running environment

Traditional coordination mechanisms are mainly considered the internal coordination of single enterprise which cope with uncertain market under fixed environment. Therefore the coordination mechanisms survive in a "steady" environment. Under the environment of e-business, information sharing between enterprises of supply chain is more frequent. In order to reply the changeable market quickly, rational coordination mechanisms must be set up between enterprises of supply chain. However, the change of decision for any enterprise will cause the difference of decision for other enterprises because of interdependence of supply chain enterprise. The coordination mechanisms must be fit for dynamic environment. The coordination mechanisms should be made by considering the uncertain factors of environment of supply chain management, so that the supply chain can deal with the market quickly and

effectively.

2) Decision mode

There are two kinds of decision modes for the coordination mechanisms of traditional supply chain: centralized decision and distributed decision. Centralized decision supposes that the decision center can get information of all relevant enterprises or departments, and make relevant decision based on all information. Distributed decision is another kind of extreme form of decision mode, and the decisions are made by each enterprise or department according to their own information. Under the environment of e-business, depending on information technology, information can be transmitted fast among enterprises, the operation of synchronization among enterprises will be realized, and the purpose of responding to market and meeting customer's demand will be achieved. But the real supply chain is often a competitive-cooperative system formed by different enterprises whose purpose is to maximize their own interests, so the key enterprise can not obtain all information needed. While setting up the coordination mechanisms, enterprises should consider both their own existing information and the information of partners that can be received, which can expand the whole optimization space of supply chain.

3) Feedback mechanism

The implementation of any system must set up a feedback mechanism of information. The implementing quality of coordination mechanism is strongly connected with the feedback mechanism. The traditional feedback mechanism is a type of chain. The feedback of information is transmitted from one department to another, which forms a parallel mode of information transmission with organizational structure. Supply chain based on e-business has a feedback mechanism having the characteristics of network. The information of enterprises is transmitted not according to the organizational structure, but the different enterprises in the supply chain. Because the organizational forms of enterprises in the supply chain are different, the forms of information flow are different.

4) Transparent degree

The enterprise of traditional supply chain which regards their own interest as the center neglects the information and interests of partners. The coordination method adopted is mainly to optimize one's own interests. Under the environment of e-business, there is very strong cooperative relation between enterprises. The information of all enterprises of supply should be considered while making decisions. The coordination mechanisms of enterprises should have heavier transparency in order to gain the synchronal operation in the supply chain.

3 COORDINATION MECHANISMS BASED ON E-BUSINESS

Under the environment of e-business, information sharing between enterprises of supply chain management is increasing, and relation between enterprises is different. For example, ShiHua Ma divides the partner's type into four kinds from both increment function and competitiveness: strategic cooperative partner, competitive / technical cooperative partner, powerful cooperative partner, and ordinary cooperative partner. Kim Langfield-Smith divides it into three kinds: non-cooperation, short-term cooperation, close cooperation and cooperative network relation. So in order to stimulate partners effectively, different coordination mechanisms should be applied to different partners. The paper presents the suitable coordination mechanisms under e-business environment according to market relation and competitive-cooperative relation.

3.1 Market Coordination Mechanisms

Market coordination is used to conduct a business transaction, to purchase something at a given price. Frequently, market coordination is used to foster price competition among potential suppliers. Therefore, market transactions are often one-time transactions because a different supplier can be selected for each transaction. The information that is shared using market coordination is generally limited to the conditions of the transaction, delivery information and payment, in addition to bidding processes. Usually, market coordination will not provide a sustainable competitive advantage to a firm. Nonetheless, they tend to be simple to use and frequently offer the opportunity to purchase some items at a lower price, making them popular. We identify two market coordinative mechanisms: on-line coordination and Electronic agents.

On-line auction is a market mechanism in which the buyers and seller agree on the item to be auctioned, and on payment and delivery conditions. Under the environment of e-business, auction has new characteristics. Firstly, no relation-specific assets between buyer and seller are needed in an on-line auction on the internet. Secondly, because auctions are one-time transactions, the transaction risk is low and the operational performance risk for auctions can be high. Finally, since the preparation for and the participation in an auction is time consuming, the frequency of transaction needs to be low. Higher frequency might make other coordination mechanisms preferable.

Electronic agents that focus on studying the coordination mechanism between the entities are the core in the field of the distributed artificial intelligence in recent years. They connect commercial goal with customer's interests to reduce the transaction costs between buyers and sellers automatically. Electronic agents are part of the IT infrastructure since they are used for doing business with all possible customs or

suppliers. There are no relation-specific assets. Further, automated transactions are one-time transactions, which makes the transaction risk low. For the same reason, operational performance risk is considered medium to high.

3.2 Competitive-cooperative Coordination Mechanisms

With the violence of market competition, the relation between enterprises changes to competition and competition-cooperation. Because the foundation of coordination is cooperation, market mechanism can be adopted under competitive relation, while competitive-cooperative coordination should be applied under competitive-cooperative relation. The contents of competitive-cooperative coordination include not only price, but also time, quantity, planning, stock, etc. coordination. Implementing competitive-cooperative coordination requires medium asset specificity, which is lower than implementing an EDI link, but it is higher than implementing other market mechanisms. Transaction risk due to closeness of the relation is high given that competitive-cooperative coordination is generally implemented with key members of the supply chain. Operational performance risk is low since management may implement coordination to actively manage the relation and, ultimately, to reduce risk of operational failure. Two competitive-cooperative coordinative mechanisms are identified: stock coordination and planning coordination.

Stock coordination can improve the systematization and integration of supply chain, and strengthen the agility and response of enterprise through controlling the inventory. The emphasis of stock coordination is to reduce "bull whip effect" caused by the separating stock of enterprises and improve synchronization of supply chain. Stock coordination that reflects new-type cooperative relation of supply chain is a stock management of risk sharing

Planning coordination is a new-mode of supply chain coordination mechanism and is considered requiring high cooperative degree between enterprises. It emphasizes the information sharing that should carry on from drawing up to implementation of the plan between enterprises, which increase the whole flexibility of supply chain.

4 CONCLUSIONS

With the development of e-business, information transmission between enterprises of supply chain becomes faster and faster. At the same time, the relation between enterprises changes into not only competition but also cooperation. So supply chain management changes greatly, and as the core of supply chain management, coordination has heavy difference too. Firstly, the influences of supply chain management that

are brought into by e-business are analyzed. Secondly, the differences of coordination mechanisms between traditional environment and e-business environment are compared. Finally, the suitable coordination mechanisms under e-business environment are put forward. There are positive impetuses for the administrator to identify the suitable coordination under all kinds of business environment. Applying different coordination mechanisms to different partners can greatly improve the performance of supply chain.

ACKNOWLEDGEMENT

This paper is supported by the Natural Science Foundation of educational department of Shaanxi province.

REFERENCES

- [1] Sebastian J. Garcia-Dastugue, Douglas M. Lambert, "Internet-enabled coordination in the supply chain", *Industrial Marketing Management*, Vol. 32, No. 2, pp251-263, 2003.
- [2] Xu Bo-yi, Tang Yue, Feng Ning, "A Case Study on E-business Supply Chain Management", *proceeding of the National Youthful Council on Management Science and System Science*, Vol. 1, pp452-458, Dalian, 2001.
- [3] Lan Bo-xiong, Zheng Xiao-na, Xu Xin, "Supply chain management in E-business era", *Chinese Journal of Management Science*, Vol. 8, No. 3, pp1-7, 2000.
- [4] Zhang Wei, Huang Xiao-yuan, "Impact of E-business on supply chain integration", *Journal of Northeastern University(Social science)*, Vol. 5, No. 4, 2003.
- [5] Richard Lancionia, Hope Jensen Schaub, Michael F. Smithc, "Internet impacts on supply chain management", *Industrial Marketing Management*, Vol. 32, No. 2, pp173-175, 2003.
- [6] Zillur Rahman, "Internet-based supply chain management: using the internet to revolutionize your business", *Information Management*, Vol. 23, No. 2, 2003.
- [7] Christoph Schneeweiss, Kirstin Zimmer, "Hierarchical coordination mechanisms within the supply chain", *European Journal of Operational Research*, Vol. 153, No. 4, pp687-703, 2004.
- [8] Kirstin Zimmer, "Supply chain coordination with uncertain just-in-time delivery", *INT.J. Production Economics*, Vol. 77, No. 1, pp1-15, 2002

Decision Model for COTS Component Procurement Based on Case-based Retrieval and Goal Programming

Xinwei Yuan¹, Jingbing Yi¹, Jun Cai²

¹ School of Management, Xi'an Jiaotong University, Xi'an 710049, China

² Shanghai Sugar Cigarette & Wine co. LTD, Shanghai 200001, China
yuanxinwei1975@sina.com

ABSTRACT

Compared with traditional information system development methodology, COTS-based information system has the following advantages: Avoid expensive development and maintenance; frequent upgrades often anticipate organization's need; rich functionality; mature technologies; tracks technology trends, etc. However, how to select appropriate COTS components is a complex problem. For improving the accuracy of decision-making in COTS component procurement, a two-period model is put forward. In the first period, the procurement requirement of each COTS component is compared with a COTS component case base by case-based retrieval (CBR) and the initial candidates are selected. In the second period, a (0-1) integer goal programming model is created to optimize cost and time of the whole COTS-based system, and help decision makers to decide the final candidates. Case shows that the two-period method declines the complexity of computation and increases the rationality of decision.

Keywords: COTS (Commercial Off-The-Shelf), component procurement, case-based retrieval, goal programming

0 INTRODUCTION AND BACKGROUND

COTS (Commercial Off-The-Shelf) component is a kind of software product that has the following characteristics [1]:

- ò Sold, leased, or licensed to the general public;
- ò Offered by a vendor trying to profit from it;
- ò Supported and evolved by the vendor, who retains the intellectual property rights;
- ò Available in multiple, identical copies;
- ò Used without source code modification.

Compared with traditional information system development methodology, COTS-based solution has the following advantages: Avoid expensive development and maintenance; frequent upgrades often anticipate organization's need; rich functionality; mature technologies; tracks technology trends, etc [2].

A lot of organizations and scholars have done a series of research on COTS component in recent years. The European Union's Framework V programme has funded a series of new research initiatives in component-based software engineering (CBSE) [3]. Gardler et al. [4] research how to support component-based software evolution at strategic level of e-Business. Maiden et al. [5] discuss how to acquire COTS software selection requirements. Some scholars give guidance about COTS component evaluation and selection [6,7,8]. Sundarraj et al. [9] create a multi-period optimization model for the procurement of component-based enterprise information technologies.

The optimization model for the procurement of COTS component given by Sundarraj considers both component procurement cost and inter-component integration cost. That makes decision more scientific.

However, the model cannot explain how to deal with the attributes difficult to be described quantitatively. In this paper, the kind of attribute is considered and a two-period decision model for COTS component procurement is given. In the first period, initial candidates are selected by case-based retrieval (CBR). In this period, the procurement requirement of each COTS component is viewed as a new problem and is compared with a COTS component case base. Of course, a COTS component case base should be created in advance by collecting information from component marketplace and scoring attributes for each COTS component. In the second period, final candidates are selected by (0-1) integer goal programming from initial candidates.

1 COTS-BASED SOFTWARE SYSTEM

A COTS-based software system means that the software system is developed by purchasing and integrating COTS components. Compared with traditional system, a COTS-based software system has longer analysis and design phases, whilst the implementation phase focuses on COTS component integration rather than on programming [4]. In implementation phase, the main task is inter-component integration. The glue code programming averagely is 37% of effort required to complete various COTS-based development activities [10].

A process model [4] for developing COTS-based e-Business system is shown in Figure 1.

E-Business Maturity Model (EBMM) is a strategic positioning and planning tool. Pattern library stores software patterns mapping business patterns that represent e-business "best-practice" processes. In analysis and design phase, EBMM is applied to export

business patterns mapping business strategy and pattern library is used to export software patterns mapping business patterns. After analyzing the software patterns, component procurement requirements are acquired. In

implementation phase, according to requirements, components are tested and purchased, and then, system is developed by components integration.

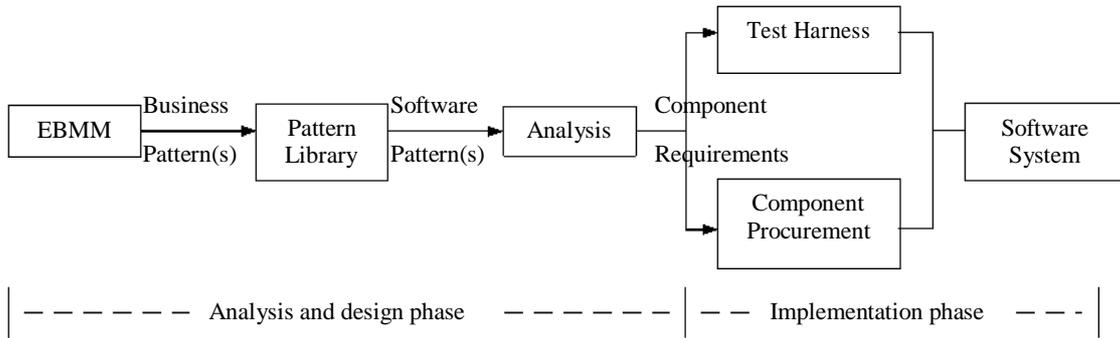


Figure 1 Process model for developing COTS-based e-Business system

In this paper, the work mainly focuses on the component procurement decision activity in implementation phase. On the basis of component requirements, CBR and goal programming are applied to create a two-period decision model.

The details about the model are given in Section 2.

2 MODEL

COTS component procurement decision-making is a complex process. In the process, large numbers of attributes should be considered comprehensively. Some attributes are shown in table 1.

Table 1 The attribute of COTS component

Category	Attribute
Technical	Application domain
	Architecture level
	Architecture compatibility
Quality	Functionality
	Reliability
	Usability
	Efficiency
	Maintainability
	Portability
	Effectiveness
	Productivity
	Safety
	Satisfaction
Vendor	Capability maturity level
	Market share
	Market size
	Market stability
	Financial viability
	Management viability
	R&D viability
	Product support
	Vendor reputation
	Training
	Delivery time
	License type
Cost	Acquisition cost
	Ownership cost

According to whether easy to be described quantitatively, attributes are categorized into two groups: quantitative ones and non-quantitative ones. Quantitative attributes, such as cost and time, satisfy additivity principle and are easy to get solution by goal programming. Non-quantitative attributes, such as functionality, are non-additivity and not easy to be applied to goal programming. In this paper, CBR is used to aggregate the non-quantitative attributes.

In section 2.1, the method how CBR is applied to candidate selection is introduced in detail. In section 2.2, the goal programming model is given.

2.1 CBR(Case-Based Retrieval)

Nowadays, there is no uniform and accepted approach for COTS component procurement decision-making. We find that selecting components from marketplace is similar to retrieving components from case base in principle. So, CBR is applied and a nearest neighbor retrieval function is given as formula (1) shows[11] :

$$\frac{\sum_{i=1}^n W_i \times \text{sim}(f_i^I, f_i^R)}{\sum_{i=1}^n W_i} \tag{1}$$

Where W_i is the weight of attribute i , $\text{sim}()$ is similarity function, and f_i^I and f_i^R are the values for the attributes of the input and retrieved cases. The retrieved case with a higher aggregate match score represents the nearer match.

Initial candidates are selected from COTS marketplace according to the threshold β defined in advance.

2.2 Goal Programming

Integer goal programming is applied to select the final candidates from initial candidates. Two main quantitative

objectives, cost and time, are considered.

3 EXAMPLE

The total cost of a COTS component is made up of acquisition cost and ownership cost. Acquisition cost includes price and procurement expense. Ownership cost mainly includes inter-component integration cost.

Assume that company *H* has three COTS component requirement areas: *A*, *B*, and *C*. The inter-component integration relationship is *A*→*B*→*C*. Namely, *A* links with *B* and *B* links with *C*. The candidate set of *A* is {*A1*, *A2*, *A3*}. The candidate set of *B* is {*B1*, *B2*, *B3*}. The candidate set of *C* is {*C1*, *C2*, *C3*}.

The total Time is composed of delivery time and inter-component integration time.

In the first period, initial candidates are selected by CBR.

The (0-1) integer goal programming model is given as follows:

The similarity function is defined as formula (2).

$$\text{Min } w_1 v_1 + w_2 v_2$$

$$\text{s.t.}$$

$$\text{sim}(f_i^L, f_i^R) = 1 - \frac{|f_i^L - f_i^R|}{f_i^L + f_i^R} \in [0, 1] \quad (2)$$

$$\sum_{j=1}^{J_i} \sum_{i=1}^I (p_i^j + e_i^j) x_i^j + \frac{1}{2} \sum_{j \neq j' \neq j''} c_{i' i''}^{j j'} x_{i'}^{j'} x_{i''}^{j''} - v_1 = c_{min}$$

Because requirement areas are viewed as frame of reference to score, we can assume the values for attributes of each requirement area have been standardized, namely, $f_i^L \equiv 1$. Then, formula (2) can be simplified as formula (3).

$$\sum_{j=1}^{J_i} \sum_{i=1}^I t_i^j x_i^j + \frac{1}{2} \sum_{j \neq j' \neq j''} t_{i' i''}^{j j'} x_{i'}^{j'} x_{i''}^{j''} - v_2 = t_{min}$$

$$\text{sim}(f_i^L, f_i^R) = f_i^R \quad (3)$$

$$\sum_{j=1}^{J_i} x_i^j = 1 \text{ for all } i \in I$$

The amount of attributes is large in practice. For simplifying computation, only four attributes—functionality, reliability, safety, product support degree—are considered in this example. In table 2, the values of attributes are given by score method.

$$\sum_{j=1}^{J_i} x_{i'}^{j'} = 1 \text{ for all } i' \in I$$

$$\sum_{j=1}^{J_i} x_{i''}^{j''} = 1 \text{ for all } i'' \in I$$

$$x_i^j, x_{i'}^{j'}, x_{i''}^{j''} \in \{0, 1\}$$

$$v_1, v_2 \geq 0$$

$$w_1, w_2 \geq 0$$

Table 2 The values for attributes

	Functional -ity	Reliability	Safety	Product support degree
A1	1.0	0.9	0.9	0.9
A2	0.9	0.8	0.8	0.8
A3	0.8	0.8	0.7	0.6
B1	0.8	0.9	0.8	0.9
B2	0.9	0.9	0.9	0.9
B3	0.7	0.6	0.7	0.8
C1	0.8	0.8	0.9	0.9
C2	0.9	1.0	0.9	0.9
C3	0.9	0.8	0.9	0.8

Notation: p_i^j is the price of candidate component M_i^j ; e_i^j is the procurement expense of component M_i^j ; $c_{i' i''}^{j j'}$ is the integration cost of component $M_{i'}^{j'}$ and component $M_{i''}^{j''}$; t_i^j is the delivery time of component M_i^j ; $t_{i' i''}^{j j'}$ is the integration time of component $M_{i'}^{j'}$ and component $M_{i''}^{j''}$; $i, i', i'' \in I$; I is the set of component requirement areas that needed to be covered; $j, j', j'' \in J$; J_i is the set of candidate components for component requirement area i ($i \in I$);

The weights of four attributes are 0.4, 0.2, 0.2, and 0.2 in turn. The threshold β is 0.8. Input the data above into formula (1) and formula (3). The result shows that *A3* and *B3* should be removed from candidates. Therefore, *A* has two choices {*A1*, *A2*}; *B* has two choices {*B1*, *B2*}; and *C* still has three choices {*C1*, *C2*, *C3*}.

$$x_i^j, x_{i'}^{j'}, x_{i''}^{j''} = \begin{cases} 1, & \text{if } M_i^j, M_{i'}^{j'}, M_{i''}^{j''} \text{ is selected} \\ 0, & \text{otherwise} \end{cases};$$

In the second period, final candidates are selected by (0-1) integer goal programming. The data in tables below is given by Delphi method.

v_1, v_2 is the eccentricity coefficient; w_1, w_2 is the weight of v_1, v_2 ; c_{min} is the limitation of total cost; t_{min} is the limitation of total time.

Table 3 Inter-component integration cost (hundred \$)

	A1	A2	B1	B2	C1	C2	C3
Price (hundred \$)	65	75	20	18	45	40	43
Procurement expense (hundred \$)	6.5	7.5	2	1.8	4.5	4	4.3
Delivery time (hour)	10	15	18	12	6	9	12

Table 4 Inter-component integration time (hour)

	A1	A2	B1	B2	C1	C2	C3
A1	0	0	3	2	0	0	0
A2	0	0	2.5	2.8	0	0	0
B1	3	2.5	0	0	2	1	2
B2	3	2.8	0	0	2	1.6	1.5
C1	0	0	2	2	0	0	0
C2	0	0	1	1.6	0	0	0
C3	0	0	2	1.5	0	0	0

Table 5 Price, procurement expense, and delivery time

	A1	A2	B1	B2	C1	C2	C3
A1	0	0	1	2	0	0	0
A2	0	0	1.5	0.8	0	0	0
B1	1	1.5	0	0	1	1.5	2
B2	2	0.8	0	0	1	0.8	1.5
C1	0	0	1	1	0	0	0
C2	0	0	1.5	0.8	0	0	0
C3	0	0	2	1.5	0	0	0

$$w_1=0.75, w_2=0.25, c_{min}=130, t_{min}=30.$$

Input data above into goal programming model and the result is $v_1=8.9$, $v_2=3.8$, $w_1v_1+w_2v_2=7.625$. The final candidates—A1, B2, and C2—are purchased.

4 CONCLUSIONS AND EXTENSION

According to different features, the attributes of COTS component are categorized into quantitative ones and non-quantitative ones. To non-quantitative attributes, CBR is applied to aggregate the evaluation values. To quantitative attributes, goal programming is applied to decide the final candidates. The two-period method declines the complexity of computation and increases the rationality of decision. A series of extensions are possible with the work described in this paper. Firstly, the collection of data may be more scientific and the sensitivity of parameters should be analyzed. Secondly, for a large-scale system, the computing would be

complex and efficient solution algorithms should be developed. Finally, the issues about decision support system aiding COTS component procurement decision-making are worth to be researched more deeply.

REFERENCES

- [1] Brownsword, L., Oberndorf, T., Sledge, C.A., "Developing new processes for COTS-based systems", *IEEE Software*, No. 7, pp48-55, 2000.
- [2] Boehm, B., Abts, C., "COTS integration: Plug and Pray?", *IEEE Computer*, Vol. 32, No. 1, pp135-138, 1999.
- [3] Maiden, N.A., Kim, H., Ncube, C., "Rethinking process guidance for selecting software components", *Proceeding of the first International Conference on COTS-Based Software Systems*, No. 2, pp151-164, 2002.
- [4] Gardler, R., Mehandjiev N., "Supporting component-based software evolution", *Lecture Notes in Computer Science*, Vol. 2591, No. 2, pp103-120, 2003.
- [5] Maiden, N.A., Ncube, C., "Acquiring COTS software selection requirements", *IEEE Software*, Vol. 15, No. 2, pp46-56, 1998.
- [6] Lawlis, P.K., Mark, K.E., Thomas, D.A. etc. "A formal process for evaluating COTS software products", *IEEE Computer*, No. 5, pp58-63, 2001.
- [7] Kontio, J., "A case study in applying a systematic method for COTS selection", *IEEE Proceedings of 18th International Conference on Software Engineering(ICSE)*, pp201-209, 1996.
- [8] Kunda, D., "STACE: social technical approach to COTS software evaluation", *Lecture Notes in Computer Science*, Vol. 2693, pp64-84, 2003.
- [9] Sundarraj, R.P., Talluri S., "A multi-period optimization model for the procurement of component-based enterprise information technologies", *European Journal of Operational Research*, No. 146, pp339-351, 2003.
- [10] Basili, V.R., Boehm B., "COTS-based systems top 10 list", *IEEE Computer*, Vol. 34, No. 5, pp91-95, 2001.
- [11] Kolodner, J., *Case-based reasoning*, San Mateo, CA: Morgan Kaufmann, 1993.

Design of an E-SCM system for Auto-Parts Companies

Junsu Lee, Jaebyoung Rhu, Kiseong Kim, Myunghyun Oh, Jongyeong Lee, Byungho Jeong
Dept. of I&IS Eng., Chonbuk University, Chonju, Chonbuk, 561-756, Korea

ABSTRACTS

This paper deals with the design of e-SCM for a supply chain of a commercial vehicle company. Two first vendors and two second vendors of the company are included in the objects of developing e-SCM system. We analyzed the information flow and business process between 1st vendor and 2nd vendors. The database (DB) has to be designed for contents of easy access, correction and update, because DB is indispensable for developing the new system. A relational database concept was used to design tables which maintain data needed in the system. The paper uses data flow diagram to describe TO-BE model of the e-SCM system.

Keywords: e-SCM, SCM, VMI

1. INTRODUCTION

Intensive competition in recent global markets, the introduction of products with short life cycles and the heightened expectations of customers have forced business companies to focus their attention on their supply chain. In a typical supply chain, raw materials are purchased, items are produced at one or more factories, shipped to warehouses for intermediate storage, and then shipped to retailers or customers. In this context, supply chain management(SCM) is a set of approaches utilized to efficiently integrate suppliers, manufacturers, warehouses, and stores, so that an end item is produced and distributed at the right quantities, to the right location, at the right time, in order to minimize system wide costs while satisfying service level requirements [1, 5]. The supply chain of a commercial vehicle company has more complex procurement chain. It is most important to manage the procurement process of auto

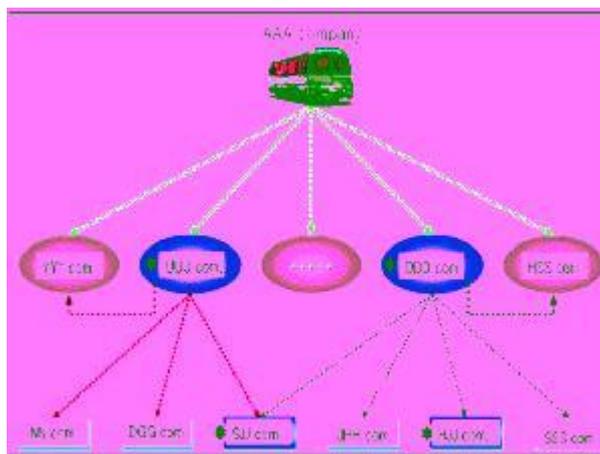
parts in the supply chain of a commercial vehicle company. Thus this paper deals with the design of an e-SCM system for small and medium size auto-part companies that are the first and second vendors of a commercial vehicle company.

The SCM system has to be designed in integrated point of view to be used successfully [2, 3]. Even though every researcher defines the integration differently, in general, supplier integration, alignment of purchaser and supplier, coalition between suppliers, and supply chain synchronization are included in the concept of integration [4]. In those senses, the supply chain of a commercial vehicle industry is more complex than those of other industries and especially, a procurement supply chain is relatively important in the industry. However, from the characteristics of domestic industrial environment, the relations with a vehicle company and its vendors forms a hierarchical supply chain structure with dependency rather than cooperative independency. Since the competitive power of a vehicle industry depends on the competitive power of whole supply chain rather than that of the industry itself, we need the effective methods to manage the whole supply chain

including the procurement supply chain reasonably. Thus, this paper dealt with analysis of business process and information flow, system design for developing e-SCM system of some auto-part companies. Section 2 provides a brief overview of sample supply chain. Section 3 describes the results of AS-IS analysis to understand intra-company business process and inter-company data flows. The TO-BE model is given for design of e-SCM system based on previous results in section 4.

2. SELECTED PROCUREMENT SUPPLY CHAIN FOR DEVELOPMENT OF AN e-SCM SYSTEM

Four auto part companies are selected to develop the e-SCM system in this study. Two of them are the first vendors and others are the second vendors of company 'AAA' which assembles trucks and buses. A company 'UUU' is the first vendor and has about 50 employees.



*: companies to be considered in this research

< Figure 1> A sample supply chain for procurement of the vehicle company

The roof of buses is main item which the company 'UUU' produces and supply for the company 'AAA'. Some of other first and second vendors of the company 'AAA' furnish the company 'UUU' sub-parts that are

required to make the roof of buses. A company 'DDD' is also one of the first vendors of the commercial vehicle company 'AAA' and about 60 employees work for this company. The main products are a frame of buses and bumpers of trucks. The sub-parts for frames and bumpers are made for itself or are supplied from other first and second vendors of 'AAA'. The second vendors 'SJJ' and 'HJJ' are included in the procurement supply chain of this study. <Figure 1> shows the part of the supply chain for this study.

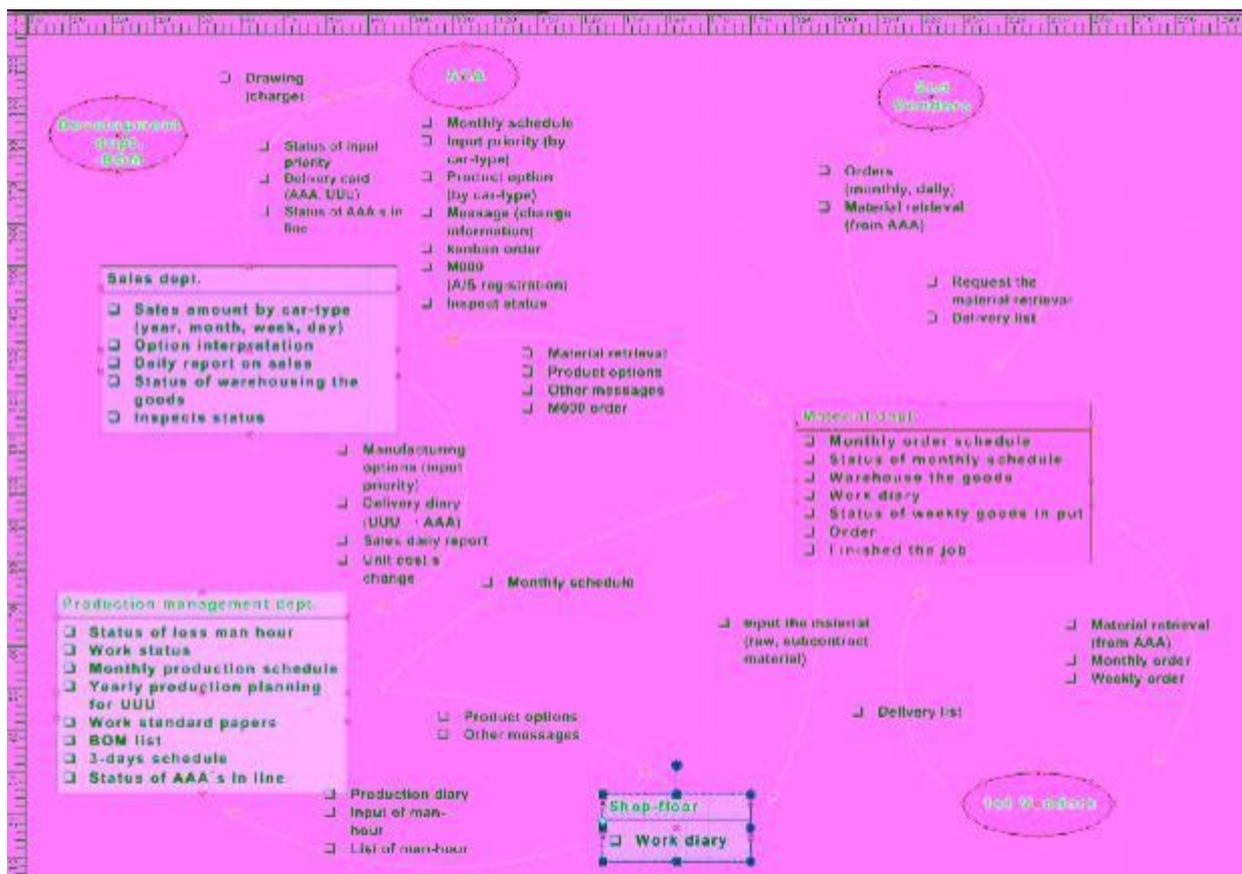
3. SYSTEM ANALYSIS

3.1 Information flow within procurement process

We analyzed a delivery process that the company AAA transmits to the 1st vendors' monthly production plan, the input sequence of the final assembly line and specification reports. The sales part of the 1st vendor gets these order information through a VAN system and passes those information on the other part, such as production control and procurement department of the company. The operation manager makes a monthly production plan from AAA's monthly plan and a 3-days production schedule on the basis of AAA's sequence

information and specification report. Currently, the monthly plan and 3-days schedule made by the operation manager are notified to related divisions (purchase section, production section, and sales section) and related 2nd vendors in a document. The purchasing manager calculates net requirement of each part on the basis of BOM (Bill of Material) and inventory records of the parts. And then purchasing orders are delivered to related 2nd vendors.

The raw materials are requested and released to produce bus roof by 3-days production schedule. The released materials are transmitted to the process line, and then the results of production are recorded in job card and work diary to the end of manufacturing. A delivery man of company UUU or DDD distributes completed parts with delivery card to the assembly line of company AAA according to a daily delivery schedule. The sales manager checks the daily delivery results from AAA's VAN system. Even though business processes of two companies UUU and DDD are slightly different each other, the principal business processes of two companies are almost same. <Figure 2> shows the work flow of two companies.

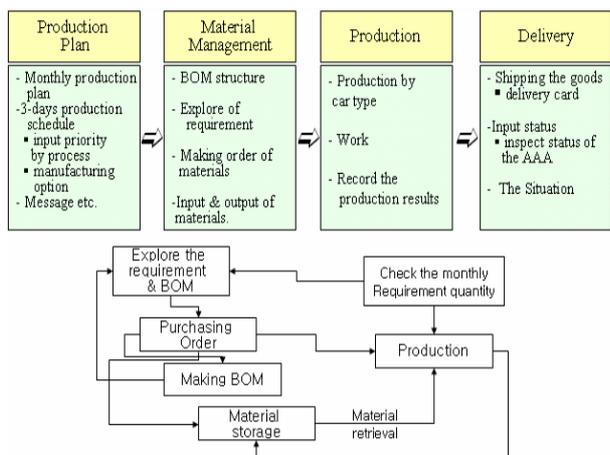


3.2 Analysis of the work flow

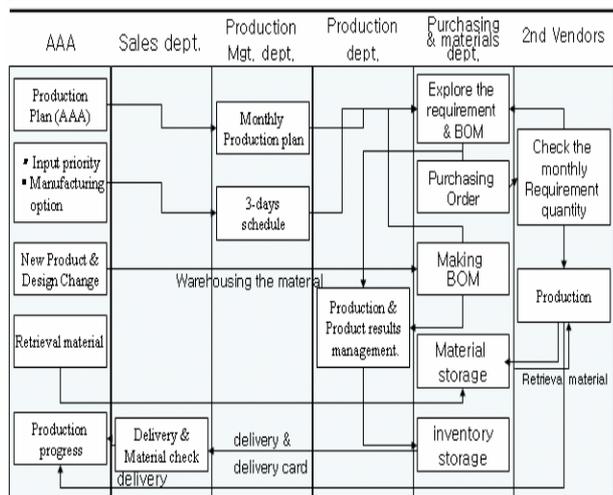
The sales department receives the monthly production plan, the input sequence of each assembly line, manufacturing specification and the other special

messages from AAA's VAN system and distributes to the department of production management and purchasing department. The AS-IS analysis about information flow between departments of the 1st vendors was accomplished for main four departments (sales, production, purchasing

materials and production management). The sales department supplies materials, which are received from AAA, to the production department. The department delivers the completed parts into the corresponding assembly line of company AAA and checks the production line stands of AAA and delivery status etc. The department of production management makes each production plan(monthly and 3 daily. The purchasing department calculates the material requirements by a monthly and 3-days schedule and makes an order to the 1st and 2nd vendors. <Figure 3> summarizes the work flow related with production management and material management. The whole business flows including the 2nd vendors are shown in <Figure 4>.



<Figure 3> Work flow of production management and material handling



<Figure 4> Work flow for 1st vendors and their divisions, AAA company and 2nd vendors.

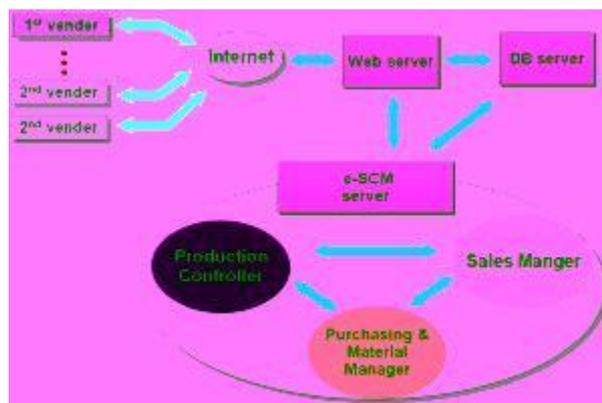
4. SYSTEM DESIGN

4.1 Conceptual Design of an e-SCM [7]

The system was designed based on client/server structure which is a typical structure of a web-based system. The conceptual framework of the system is shown in <Figure 5>. The figure shows that the system consists of a web

server, a DB server, and three functional modules such as production controller, sales manager, and purchasing & material manager. The users of the 1st and 2nd vendors access to web-server through Internet and the web-server makes connection to an appropriate module.

The second vendors can check out the net requirement and current inventory records from the web page of the 1st vendor any time, and decide how many or what items they should deliver. To do this, the 1st and 2nd vendors must agree with the proper inventory level or average inventory level for each part. That is, vendor managed of the 1st vendors.



<Figure 5> A Conceptual framework of the e-SCM system.

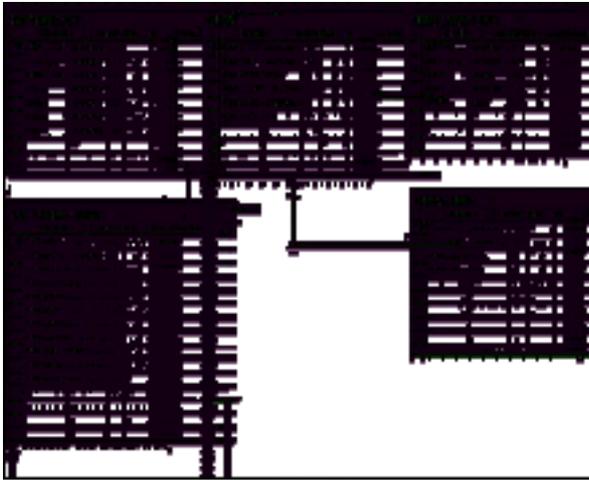
4.2 Design of Database

As DB is indispensable for developing the e-SCM system, it has to be designed for contents of easy access, correction and update[6]. DB includes valuable and various data about primary data of manufacturing item, parts, sub vendors, drawing data, inventory records, production schedule, bill of material information and production or delivery data. DB takes different effect for performance by times of access. Therefore, DB, in this study is designed with divided table by size of data and times of access to table with planning system. A relational database concept was used to design the DB which maintains data needed in the system.

<Figure 6> shows the relations with some tables and the table ASSY which stores primary data about finished goods. ASSY table includes basic information about finished part such as product ID code, price, car-type etc. ASSY_INV maintains product ID code and on-hand inventory of finished item at company AAA and the 1st vendor. ASSY_SKETCH table consists of the name and code of drawing, the version of drawing and the file name of drawing etc.

Part DB includes various kinds of primary information such as drawing data, inventory information and manufacturing style etc. Since the frequency of access varies according to kinds of data, DB table was partitioned by the frequency and size of data. PART DB

stores basic information about each part including part ID code, part name, packaging unit, order method, type of classification, and the average monthly requirement etc. PRAT_INV has on-hand inventory, amount of anticipated requirement at 1, 2 and 3 days after. The PART tables and their relation are shown in <Figure 7>.



<Figure 6> Relation Diagram of tables related with finished item of the 1st vendor

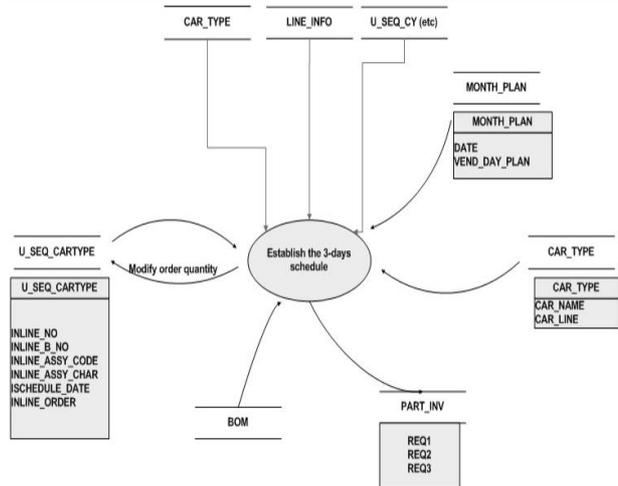


<Figure 7> Relation Diagram related with PART DB

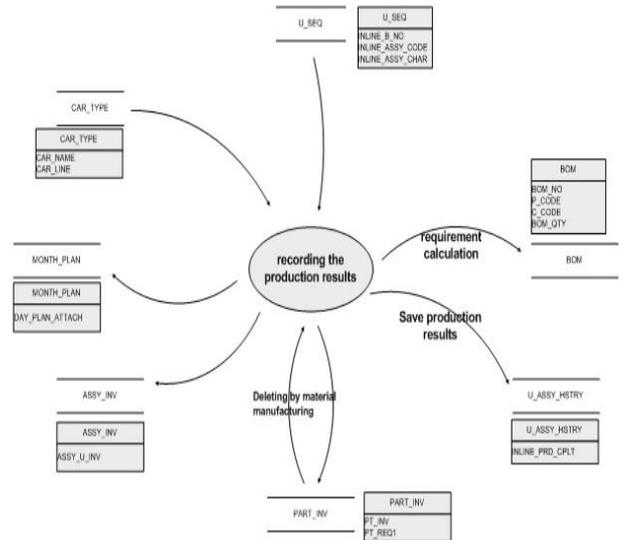
4.3 Data Flow Diagram

This section describes data flow for some fundamental functions using data flow diagram (DFD). <Figure 8> summarizes data flow between DB tables when the system makes a 3 days schedule. The system has to reference the input sequence of company AAA, car type information, monthly production plan, BOM data, and inventory data for materials etc. Whenever a 3 days schedule is made, part quantity, which is required at 1, 2 and 3 days after, is revised based on the updated 3 days input sequence. Workers put parts or material into the appropriate line as the input sequence of 3-days schedule. After completion of production, the quantity of completed production is recorded for later use. When the completed jobs are input into the system, inventory level

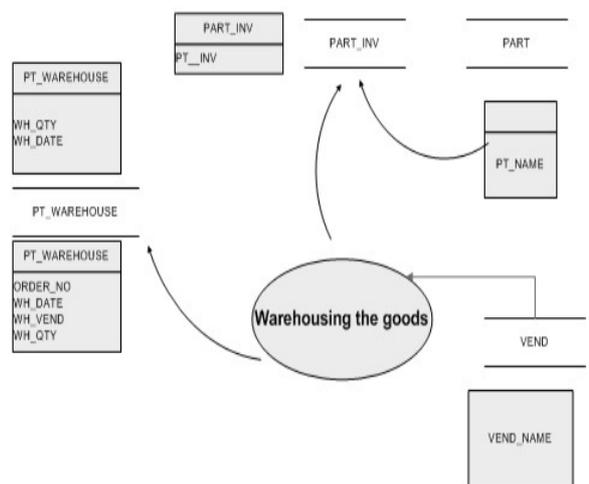
of parts, which are used to produce them, is reduced by BOM propagation. <Figure 9> shows a DFD to do this. A data flow diagram for stocking purchased parts or material is shown in <Figure 10>.



<Figure 8> Data Flow Diagram for 3-days schedule



<Figure 9> DFD for recording completed production



<Figure 10> DFD of warehousing registration

5. CONCLUSION

This paper is a case study to develop an e-SCM system to control a part of procurement supply chain of a commercial vehicle company. To do this, we analyzed the business process, information flow, and material flow among companies in the selected chain. And then, we designed the e-SCM system based on the results of the analysis. We designed the system to be developed based on web technology. The monthly plan and 3-days schedule of the 1st vendor will be made using monthly plan and input sequence of AAA. The second vendors can check out the net requirement and current inventory records of parts through the web page of the 1st vendor any time, and decide how many or what items they have to deliver. That is, vendor managed inventory(VMI) concept was applied to control inventory between the 1st and 2nd vendors.

ACKNOWLEDGEMENT

This research was supported by the Program for the Training of Graduate Students in Regional Innovation

which was conducted by the Ministry of Commerce, Industry and Energy of the Korean Government.

REFERENCES

- [1] Hukusima Yoshiyaki, *SCM management revolution*, 21st century books, 1998
- [2] Johnson, R. & Lawerence, P., "Beyond vertical integration-the rise of value adding partnership", *Harvard Business Review*, pp. 94-101, July-August, 1998
- [3] Walton, S. V. & Gupta, N.D., "Electronic data interchange for process change in an integrated supply chain", *International Journal of Operations & Production Management*, Vol. 19, pp. 45-73, 1999
- [4] Korea Management Association consulting, *Production system innovation, Synchronization production system*, 1990
- [5] Mentzer, J.T., W. DeWIT t., J.S. Keebler, S.H. Min, N.W. Nix, SmlT h and Z.G. Zacharia, "Defining Supply Chain Management", *Journal of Business Logistics*, Vol.22, No.2, pp. 4~6, 2001
- [6] Cho, DongIl, *Database Programming for Web applications*, Samyang media, 2003
- [7] Gary, P. Schneider, *Electronic Commerce*, Thomson, 2003

Development Stages of Supply Chain Management in E-Business Environment

Jeongwook Choi

Business School, Kookmin University, Seoul, 136-702, Korea
jwchoi@kookmin.ac.kr

ABSTRACT

The emergence of electronic Business, or e-Business, has changed the operations and contents of traditional supply chain management or SCM. e-Business not only makes the operation of SCM more effective, but also plays the role of a major driving force to induce the development of new features of SCM by destroying the rules and orders that once governed and controlled SCM in the past. Based on this understanding, this paper focuses on studying how e-Business has changed the ways SCM creates added values in the following areas: the nature and contents, steps, fields and structures of value chains comprising SCM. To this end, this paper first divides overall development of supply chains into three stages. Then, detailed analysis of the features, development processes and characters of each development stage shall be discussed to identify the factors driving the changes. In addition, the types of developments shall be identified and discussed.

Keywords: e-Business, Supply Chain Management, Development Stages

1. INTRODUCTION

How does the e-Business change Supply Chain Management? e-Business does not simply enable customers to purchase goods and services over the Internet. The new form of Business backed by the Internet technology has profoundly changed the structure and meaning of supply chains. In this regard, this paper will look at how e-Business influences supply chains and how the influence actually changes the shape of overall supply chains.

Researches on supply chains have uncovered a number of diversified findings. Despite the diversified studies on supply chains, few studies have been performed on how e-Business changes supply chains, what would be root causes and driving forces to the change, and what it would look like after the change is done. This paper aims at establishing the development theory of supply chain evolution according to e-Business environment. In addition, two main themes are touched upon in this paper. e-Business based on the Internet platform changes supply chains. In this regard, this paper firstly classifies the structures of development stages leading to the evolution of supply chains in order to study the features and characteristics of changes by development stage. Secondly, this paper will look at the types of changes and the factors causing such changes.

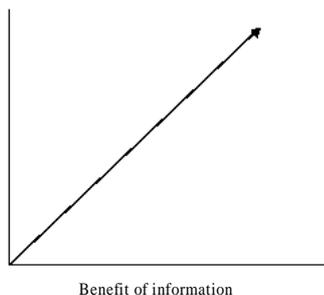
2. STAGES OF SCM DEVELOPMENT UNDER E-BUSINESS ENVIRONMENT

The advancement of e-Business has a profound impact on the development of supply chains. What are the specific impacts of e-Business on supply chain management? To answer this question, this paper looks at the influence of e-Business (or the advance of e-Business) on the development of supply chains by dividing the developments into three stages.

2.1. 1st Stage – Phase of Continuous Improvement

In principle, all developments start with a stage to improve parts which are currently unsatisfactory. Improvement refers to activities to analyze current status to come up with a better condition. Thus, an improvement is inevitably found on the continuity of existing conditions. Familiar examples of such improvements are: How can we minimize the level of inventory?; How can we accelerate the delivery of demand to customers?; How can we harmonize and match demand and supply?; And how can we enable members of supply chains to share accurate information on a real time basis and to utilize needed information in a timely manner? As the examples indicate, the stage of improvement is to address such issues as bullwhip effect, excessive inventory, mismatches between demand and supply, delayed response to customers, and silo mentality of functional areas. Such improvement is motivated by current unsatisfactory status within an organization and is performed to resolve the issues by utilizing new environment and methodologies supported by e-Business. The phase of continuous improvement can ultimately be defined as the process of eliminating wastes and non-value added works in supply chain operation. In other words, the 1st development stage is the process of achieving maximum efficiencies and effectiveness to operate supply chains.

Examples of continuous improvement stage such as to eliminate errors and redundancies in supply chain operations, to lower inventory levels, to shorten product development cycle, to eliminate bullwhip effect and so on can be found in many companies. All components of supply chains share information to maintain optimal level of inventory and to support the fastest logistic movement; and users are able to utilize analytic information required for the operations of supply chains at the time they want.



[Figure 1 : Proportional Relationship between benefit and cost of information]

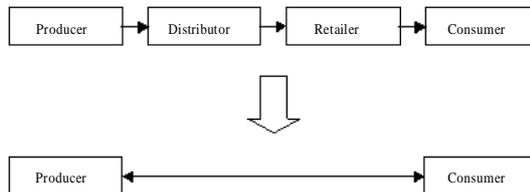
Traditionally, there has been a trade-off relationship between cost and effect of information. For instance, accurate information would have been shared through all supply chain members if value added networks or VAN was used to connect all supply chain members but this had not been realized because the costs were huge. However, the advancement of the Internet enabled automatic linkage of all members of supply chains to allow the members to exchange real-time information without accompanying large investment, as was the case in the past. At the same time, the accuracy and timeliness of information have improved greatly.

2.2. 2nd Stage – Phase of Structural Change

Customers have exerted to achieve what they wanted from traditional supply chains. The emergence of e-Business environment, however, has lifted the boundaries of thoughts of customers. In other words, once unthinkable in the traditional supply chains has turned into realities. For example, customers have become able to directly purchase goods from manufacturers without going through complicated distribution transactions. A company can act with another company located in a remote region as if they are one. These changes are not attributable to the operational efficiency intrinsic to supply chains. Rather, new environmental developments due to the emergence of e-Business have turned things that were once regarded difficult or impossible into possible features. These are not continuous or evolutionary changes from the past but disruptive or discontinuous ones. When faced with the limitations of improvement efforts that have been achieved in the 1st stage – phase of continuous improvement, companies start to recognize that the gradual improvements can no longer be used for development. Then they start to seek new ways of development or evolution. Sometimes, external shocks may transform on-going improvement activities into new types of activities or a new framework that are discontinuous from the past and out of the track of improvement projections. Both the cases of discontinuous improvement symbolize the 2nd stage – phase of structural change. In the stage of structural change, you can find:

(1) Disintermediation

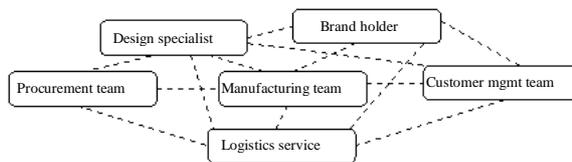
e-Business provides an environment in which customers can directly interact with producers. In other words, customer and producers do direct transactions via the Internet without utilizing traditional distribution channels.



[Figure 2 : Disintermediation]

(2) Virtual Supply Chain

Virtual supply chain refers to a system in which physically separated unit companies with unique core competencies collaborate to achieve a single target. A good example would be a collaborative shoe making network comprised of the followings: a German shoe producer which owns brand; an Italian company performs shoe designs; a Korean company takes charge of producing; a Singaporean company carries out global logistics; and a U.S. company provides customer services. As this example shows, a virtual supply network is a system in which a number of different companies gather together without regarding the limitations of time and space to act as a one entity.



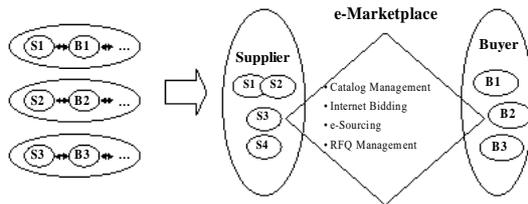
[Figure 3 : Virtual Supply Chain]

(3) e-Marketplace

Traditional relationship between suppliers and manufacturer was based on almost-fixed one-to-one ties between a supplier and a manufacturer. e-Marketplace changed ways of doing businesses of buyers and suppliers and allows them to overcome the limitations of traditional relationships by creating a single cyber marketplace where multiple suppliers do transactions with multiple buyers.

To sum up structural changes occurring in the 2nd stage: Collaboration of supply chain members has been pursued to reduce the total cycle time from customers to producers and to share information in a more efficient manner. These efforts have been discussed in the 1st stage of continuous improvement. In the 2nd stage of structural change, however, the focus of such efforts has shifted from the efficiency of a supply chain to the change of overall structure of a supply chain. In this stage, supply chains undergo discontinuous changes, not gradual improvements as are the case in the 1st stage. In this regards, we call changes occurring in this stage “structural changes” to reflect the fact that the overall framework of supply chains are subject to change.

Disintermediation, virtual supply networks or e-marketplaces are understood as types of such structural changes.



[Figure 4 : e-Marketplace]

Major driving force leading to the structural changes in the second stage would be the destruction of trade-off relationship between the richness and the reach of information where richness of information means that information contains sufficient contents to enable decision-making and reach of information means the scope within which information can be shared. For more details about the reach and richness, see Evans and Wurster (1999).

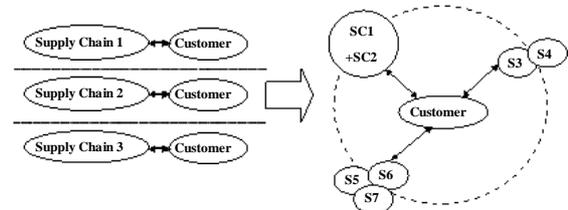
2.3. 3rd Stage – Phase of Convergence/Chaos

If the rate of structural change exceeds a certain level, such change becomes no longer controllable and starts to create spontaneous changes. This paper defines the 3rd stage as the phase of convergence/chaos changes. In this stage, the traditional definition of supply chains becomes meaningless and chaotic revolution starts to prevail. For the purpose of this paper, the chaotic revolution is defined to be a stage in which new types of converged supply chains emerge and all industries are combined with other industries, followed by the creation of new types of industries.

Values once provided by a single industry for customers come to lose their competitiveness. Thus, industries start to converge with new industries to overcome the borders of traditional industries and to provide comprehensive values to customers. For instance, on-line stock trading is a financial industry combined with IT industry. Similar examples are: mega-markets supporting purchasers to buy various products of various industries in one marketplace; integrated and converged financial service markets which have seen segmented among banks, non-banking financial institutions, insurance companies, credit union; and also the crossover competition and cooperation between credit card companies and cell phone carriers caused by the addition of credit card functional service provided by cell phone terminals. These cases manifest disappearing distinctions between traditional industries.

The speed and severity of such changes further move supply chains forward to have more and more new features. Customers come to demand all their requirements to be met at one place and the demarcations among traditional industries become

indistinct. Primary activities and supporting activities of value chains become no longer separated and the boundaries dividing industries start to disappear. Thus, the one and only yard stick of supply chain is: which supply chain delivers goods or services to customers with the highest level of customer satisfaction? When these are considered, major competitiveness of traditional supply chains, the capability of creating unique values, may become meaningless.



[Figure 5 : Converged Supply Chains]

As the figure shows, industries A, B, and C secure their own customers. These supply chains, however, start to be integrated to form a new network by integrating what customers want. As a result, supply chains by industry and by companies become no longer meaningful and, ultimately, a combination of values meeting customer needs would win the competition.

The major force driving this change is the transition of information ownership from producers to consumers. Up until now, a number of people have remarked that the paradigm has shifted from a producer's market to a consumers' market. Nevertheless, industry paradigms remained the same and the new market was not a truly consumer-oriented market although consumers could exercise more influence than before. The information revolution induced by e-Business, however, allows consumers to hold power of having complete and integrated information, which in turn enables them to select goods and industries that they want, to combine or transform producers as they want. Here, such change is defined to be chaos because it is not clear which direction the change will take. The only clear thing is that a converged and integrated supply chain which ensures the maximum customer satisfaction will become a winner in the competition.

3. SCM UNDER E-BUSINESS – TYPES AND FEATURES OF DEVELOPMENT STAGES

Previous sections have viewed how traditional supply chains change under e-Business environment. This section will analyze the types and features of such changes. The table below briefly touches upon the features of each development stages.

[Table 1 : The features and characteristics of the stages in supply chain]

Development Stage	1 st stage	2 nd stage	3 rd stage
Characteristics	Continuous and gradual improvement	Structural and discontinuous change	Chaotic and unpredictable convergence
Features	Serial supply chain	Parallel and network supply chain	Integrated and converged supply chain
Change drivers	Destruction of trade-off relationship between cost and effect of information	Destruction of trade-off relationship between reach and richness of information	Transition of Information ownership from producers to consumers
Performance measures	Effectiveness and efficiency of supply chain operation	Optimal structure of supply chain design	Maximum customer satisfaction
Change pattern Predictability	High	Medium	Low

As is shown in the above table, the first stage focuses on enhancing efficiency and effectiveness of operations and decreasing non-value added activities in supply chain operations. Such development has been available because the cost of managing supply chain information is not costly as before. Because of the continuous nature of development, it is possible to predict the development of overall operations of supply chains and the level of improvement in terms of effectiveness and efficiency. Unfortunately, however, once substantial changes have been made through the first stage of continuous improvement, it becomes very difficult to gain more desirable impact although new efforts are exercised. In other words, it becomes clear that this kind of continuous improvement almost reaches the peak of the first stage's ideal status within the current structure of supply chain. Thus, it becomes inevitable to change the structure of supply chains - external factors, not internal factors - in order to create new stage of developments. Then it begins to initiate efforts to change the supply chain environment given. In the past, most external factors were given, so it was almost impossible to change. However, the emergence of e-Business features and the application of such features opened the ways for the structural changes of supply chains. In this second stage of structural changes, unlike in the first stage, supply chains started to be re-configured, and develop more innovative structural changes.

While the first stage focused on the continuous and gradual improvement of the ways of supply chain operations, the second stage has concentrated on discontinuous development in a way to innovate overall structure of supply chains. The second stage of structural change has re-designed and re-constructed supply chains in an innovative way by utilizing the features of the Internet environment. For instance, during the second stage of development, innovative change of performing direct linkage between producers and consumers emerged, if needed, by destroying the intermediary steps linking the two parties. In other words, the second stage was about re-structuring of operations of companies and supply chains in an

innovative manner through the destruction of the previous formats. In summary, full utilization of e-Business features in the Internet environment for re-designing and re-organizing new supply chains characterizes the second development stage.

Although innovative re-structuring change has been prevalent in the second stage, competition is still limited within supply chains performing similar businesses or within similar industries. As a result, winning a competition in the second stage means a success among companies or supply chains performing similar businesses or within the same industry. Further development of e-Business technologies in the third stage of convergence and chaos, however, accelerates the advance and establishment of an era of consumers. In the new era, consumers start to ignore the boundaries embedded in the traditional industries and supply chains and require the type of a company or a supply chain that best fits their needs. In this case, however, the competitiveness of supply chains improved through innovative re-structuring activities in the second stage would be meaningless unless customers select such supply chains. Thus, consumers drive the convergence and integration of industries as they want. In other words, the convergence and integration of industries are made in a way to maximize customer satisfaction. In addition, as the development is led by customers, it is very chaotic to predict the development directions in the third stage. Thus, the convergence between companies, supply chains or industries would render a brand-new industry or supply chain.

4. CONCLUSION

This paper reviews how the emergence of e-Business changes supply chains. In this regard, features of development stages are classified, actual cases of the features are identified, driving forces of each stage are studied and analyzed, and the types of possible evolutions are suggested.

e-Business first acts as a driving force for improving current status, supports information sharing on a

real-time basis, and allows paperless management through electronic document exchanges. As a result, e-Business makes supply chains more effective and efficient as a whole.

Further advancement in e-Business destroys the trade-off relationship between richness and reach of information. This causes a brand-new development happen, leading to direct linkage between customers and producers, emergence of virtual enterprises, and a new-marketplace utilizing cyber space. In this stage, e-Business stops being a stimulus to facilitate supply chain's operational efficiency but acts as an innovative driver for changing the structure and shape of overall supply chains. One-step further advancement in e-Business blurs the demarcation lines among industries and leads supply chains into the stage of convergence and chaos full of disrupted scope and contents of competition.

The development stages are triggered by unique forces driving changes in each stage and the speed of change is dependent upon the severity of the drivers. The progress of development is not gradual; rather, it shows a number of development types.

Further researches into the characteristics of each development stage are required along with the analyses of actual cases to be observed in the industries. By doing so, models representing each development stage and more in-depth and diversified analyses of actual progresses of each stage would be carried out to track actual developments and realization of changes.

REFERENCES

1. Bradley, S., J. Hausman, and R. Nolan, (1993), *Globalization, technology and competition*, Harvard University Press
2. Cooper, M. C. and L. Ellram, (1993), "Characteristics of supply chain management and the implications for purchasing and logistics strategy," *The International Journal of Logistics Management*, vol. 4, pp. 13-24.
3. Evans, P. and T. Wurster, (1999), *Blown To Bits*, Harvard Business School Press, 1999
4. Fisher, M.L., J.H. Hammond, W. Obermeyer, and A. Raman, (1994), "Making supply meet demand in uncertain world," *Harvard Business Review*, May-June 1994, pp. 83-93
5. Gattorna, J, (1998), *Strategic Supply Chain Alignment*, Gower
6. Greis, N and J. Kasarda, (1997), "Enterprise logistics in the information era," *California Management Review*, Summer 97, pp. 49-61
7. Jones, D., P. Hines and P. Rich (1997), "Lean logistics," *International Journal of Physical Distribution*, vol 27, pp. 153-164
8. Keen, P, (1998), *Using telecommunications for competitive advantage*, Ballinger Press
9. Lee, H. L., V. Padmannabhan, and S. Whang, (1997), "The bullwhip effect in supply chains," *Sloan Management Review*, vol 38, pp. 32-43
10. Leyland, F., P.R. Berthon, R.T. Watson and G.M. Zinkhan (2002), "The Internet and the birth of real consumer power," *Business Horizons*, Vol 45, pp. 7-14
11. Malone, T., J. Yates, and R. Benjamin, (1997), *Electronic markets and electronic hierarchies*, *Communications of ACM*, 30, pp. 484-497
12. Martin, J. (1996), *Cybercorp*, Amacom
13. McKinsey and CAPS, (2000), *B2B e Marketplaces*, McKinsey & Company and CASP Research
14. Simon, R.C. (2000), "The impact of Web-based Procurement on the Management of Operating Resources Supply," *The Journal of Supply Chain Management*, Winter 2000, pp. 4-13
15. Stevens, G. (1989), "Integrating the supply chain", *International Journal of Physical Distribution & Materials Management*, vol 19, pp. 45-59
16. Schwartz, E. (1997), *Webconomics*, Ellison

E-Booking Control Problem in Sea Cargo Logistics

Chengxuan Cao¹, James S. K. Ang², Hengqing Ye²

¹ The Logistics Institute - Asia Pacific, National University of Singapore, Singapore 119260

² School of Business, National University of Singapore, Singapore 117592

ABSTRACT

In this paper, we consider the e-booking control problem which includes optimal pricing and capacity allocation. In pricing model, customers may refuse bookings based on their willingness to pay the quoted price. Since customer behavior and characteristics are highly varying and not known in advance, we develop a stochastic pricing model which tracks customer behavior as well as the arrival process to maximize profit. For the multi-period capacity allocation of e-booking control problem, we present a two-stage stochastic mixed integer programming model and a heuristic algorithm. The solution to the model is found by maximizing the expected profit over the possible control decisions under the uncertainty of shipping capacity. Finally, we give numerical experiments demonstrate the efficiency of the algorithm.

Keywords: e-booking control, container shipping industry, mathematical programming

1. INTRODUCTION

Alliances and partnerships are prevailing characteristics of the container shipping industry. A brutally competitive environment, low freight rates, and the need for carriers to increase their revenue have led to joint carrier operations. Carriers have to band together in operating alliances with the objectives of improving capacity utilization and more service flexibility. Unfortunately, this consolidation has not resulted in upward pressure on shipping rates, nor has it increased profitability for the carriers involved. The high cost of capital equipment and low rates of return that characterize container shipping compel carriers seek out every opportunity to improve their own with respect to operational efficiency and performance (see [3], [5]).

Effective booking control is a key to improved revenues and reduced container shipping costs and it has become more important when the container shipping industry has widely adopted information technologies. E-booking service supports e-business strategy for container shipping industry, which aims to make business processes easier and more cost efficient. As a part of sea cargo e-commerce strategy, web-based e-booking system enables the customer to electronically book shipments and to channel other shipment requirements to the carrier, such as cargo collection, delivery requests, and information required for customs clearance. In general, e-bookings may be made up to 30 days in advance. The customers can specify the number and type of containers needed, and service time windows at the origin and destination locations. Minimally, the time window information must include the earliest time containers may be loaded at the origin, and the latest time containers should be delivered to the destination. The objective of e-booking service is to help customers enjoy the competitive advantage of the Internet and provide an efficient, flexible and fast way of doing business. Customer Service staff can also help customers with more complicated bookings that require special expertise.

The whole industry-shippers, forwarders, carriers and consignees-benefit from minimizing or eliminating duplicate data entry. It makes the shipment process more transparent and improves the accountability of service provide. This means that the monthly freight schedule booking process can be done by pressing one button. A new e-business strategy will improve customer service and efficiency for both carriers and their customers. This move to a more efficient and effective way of making cargo booking is an important part of operational level service route planning in container shipping industry.

To obtain service, customers first check the price quote for a given origin-destination service in e-booking website, and then subsequently make bookings under the quote. With the exception of certain ancillary charges, the carrier charges the customer a fixed price for transportation, and pays the transportation service providers directly out of this fee. Therefore, it is in the carrier's interest to minimize the transportation costs for most shipments. However, there must be a level of "reasonableness" in transit times, and some customers may be willing to pay a premium for faster service.

In this paper, we consider the e-booking control problem in the container shipping industry, which includes pricing and capacity allocation. The pricing and revenue management problem is formulated as nonlinear programming in which some parameters obtained from stochastic marked point process. The capacity allocation problem is formulated as two-stage stochastic mixed integer programming model which accounts for total profit in multi-period and constraints of limited shipping capacity. Efficient algorithms for the models and numerical experiments are proposed at last. The rest of the paper is organized as follows. The problem formulation and algorithm are described in Section 2. Empirical results are given in Section 3. We conclude the paper in Section 4.

2. PROBLEM FORMULATION

2.1. Pricing and Revenue Management

Consider an e-booking system for an ocean liner operator. Cargo bookings arrive at the system according to a compound Poisson process $\{X(t)\}_{t \geq 0}$. Its jump times T_1, T_2, T_3, \dots form a Poisson process with rate $\lambda > 0$, and the jump magnitudes Y_1, Y_2, Y_3, \dots are *i.i.d.* random variables with Pareto distribution. The Pareto principle states that a large income stems from a very small number of customers. It is reasonable to assume that the freight volume of a booking follows Pareto distribution (see [1]). The Pareto probability density function with shape parameter $a > 1$ and scale parameter $b > 0$ is defined as

$$f(x) = \frac{ab^a}{x^{a+1}}, \text{ for } x \geq b \quad (1)$$

Figure 1 illustrates the Pareto probability density function for different values of shape a and scale $b=40$.

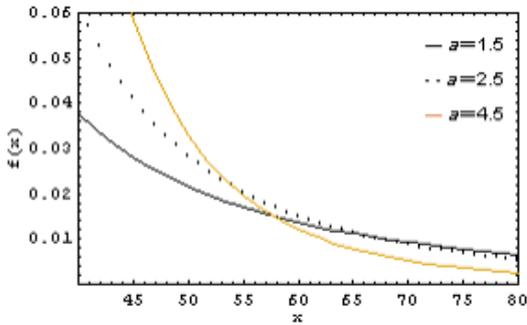


Figure 1. Pareto probability density function

When the quoted price is p , an arrival customer may accept or reject the price. Let ξ be the binary random variable, i.e., $\xi=1$ if the arrival customer accepts the quoted price; $\xi=0$, otherwise. We assume that the random variable ξ is independent of booking arrival process and the probability that an arriving customer accepts the quoted price p is a decreasing and differentiable function of p . The probability is chosen as

$$g(p) = 1 - \left(\frac{p}{\theta}\right)^\delta, \text{ for } p \leq \theta \quad (2)$$

where δ and θ are constants. By varying the parameter δ , we can make the willingness as elastic as desired. The lower the value of parameter δ , the more willing are customers to book under the quoted price. Figure 2 shows the function $g(p)$ for different values of parameter δ and $\theta=10$. The total profit in time t is:

$$Z(t) = \sum_{i=1}^{\infty} Y_i 1_{0 \leq T_i \leq t} p 1_{\xi=1} \quad (3)$$

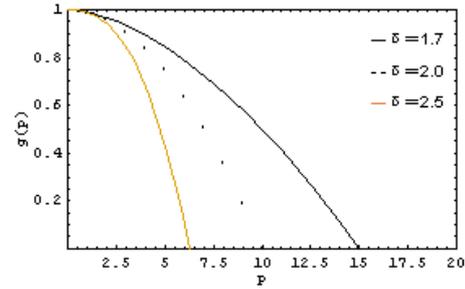


Figure 2. Probability of acceptance under quoted price p

Let total available capacity is C . Then we formulate the pricing and revenue management problem for single cargo type and single period as following nonlinear programming (M1):

$$\text{Maximize } EZ(t) = E \sum_{i=1}^{\infty} Y_i 1_{0 \leq T_i \leq t} p 1_{\xi=1} \quad (4)$$

subject to

$$E \sum_{i=1}^{\infty} Y_i 1_{0 \leq T_i \leq t} 1_{\xi=1} \leq C \quad (5)$$

Obviously, stochastic process $\{Z(t)\}_{t \geq 0}$ is a marked point process and the expectation of $Z(t)$ is (for details of marked point process, see [2]):

$$\begin{aligned} EZ(t) &= EY_1 \cdot \lambda t p \cdot P \xi = 1 \\ &= \frac{ab\lambda t p}{a-1} \left[1 - \left(\frac{p}{\theta}\right)^\delta\right] \end{aligned} \quad (6)$$

Similarly,

$$E \sum_{i=1}^{\infty} Y_i 1_{0 \leq T_i \leq t} 1_{\xi=1} = \frac{ab\lambda t}{a-1} \left[1 - \left(\frac{p}{\theta}\right)^\delta\right] \quad (7)$$

In the following theorem, we give the optimal of above nonlinear programming.

Theorem 2.1. The optimal solution of the nonlinear programming (M1) is

$$p^* = \max(\hat{p}, \bar{p}) \quad (8)$$

where

$$\hat{p} = \frac{\theta}{(\delta + 1)^{1/\delta}}, \text{ and } \bar{p} = \theta \left[1 - \frac{c(a-1)}{ab\lambda t}\right]^{1/\delta}.$$

Proof. Let

$$\eta_1(p) = \frac{ab\lambda t p}{a-1} \left[1 - \left(\frac{p}{\theta}\right)^\delta\right], \quad \eta_2(p) = \frac{ab\lambda t}{a-1} \left[1 - \left(\frac{p}{\theta}\right)^\delta\right],$$

then we easily verify that the function $\eta_1(p)$ is concave in $(0, +\infty)$ and it takes maximum at point:

$$\hat{p} = \frac{\theta}{(\delta + 1)^{1/\delta}} \quad (9)$$

On the other hand, the function $\eta_2(p)$ is decreasing in $(0, \theta)$ and it takes zero at point:

$$\bar{p} = \theta \left[1 - \frac{c(a-1)}{ab\lambda t}\right]^{1/\delta} \quad (10)$$

Therefore, the optimal solution of non-linear programming (M1) is $p^* = \max(\hat{p}, \bar{p})$.

2.2. Capacity Allocation

We assume that the capacity allocation decisions (for the entire planning horizon) have to be made, with only some knowledge of future scenarios of parameters. The overall objective is to determine a capacity allocation plan, such that the sum of each expected profit is maximized. To incorporate the uncertainty in the parameters, we assume that these parameters can be realized as one of S scenarios. The probability of scenario s will be denoted by p^s .

Before formally stating the problem, we introduce some notation:

Index Sets

- \tilde{T} : set of time periods $\{1, 2, \dots, t, \dots, T\}$.
- \tilde{K}_t : set of all cargoes received in period t , i.e. $\tilde{K}_t = \{1, 2, \dots, k, \dots, K\}$.
- \tilde{S} : set of all scenarios $\{1, 2, \dots, s, \dots, S\}$.

Deterministic Parameters

- r_{tkd} : per volume profit of cargo k which is received in period t and delivered in period d . It can be interpreted as the per volume net profit of cargo k , i.e., per volume profit of cargo k minus its per volume inventory cost.
- τ_k : due date of cargo k . Each cargo has its due date requested by shipper in its booking status.
- $v_{tk\tau_k}$: the volume of cargo k received in period t ready for delivery before its due date τ_k .
- $w_{tk\tau_k}$: weight of cargo k received in period t ready for delivery before its due date τ_k .

Random Data

- V_t^s : total available volume capacity in period t under scenario s .
- W_t^s : maximum allowable weight capacity in period t under scenario s .
- q_{d1}^s : overage cost per unit overage of volume capacity in period d under scenario s .
- q_{d2}^s : shortage cost per unit short of volume capacity in period d under scenario s .
- q_{d3}^s : overage cost per unit overage of weight capacity in period d under scenario s .
- q_{d4}^s : shortage cost per unit short of weight capacity in period d under scenario s .

Decision Variables

- x_{tkd} : binary variable, i.e., $x_{tkd} = 1$ if cargo k is received in period t and is ready for delivery in period d before its due date τ_k , 0, otherwise.
- y_{d1} : amount of overage of available volume capacity in period d .
- y_{d2} : amount of shortage of available volume capacity in period d .
- y_{d3} : amount of overage of allowable weight capacity in period d .
- y_{d4} : amount of shortage of allowable weight capacity in period d .

The multi-period capacity allocation problem can then be formulated as following two-stage stochastic mixed integer programming model (M2):

$$\text{Maximize } z = \sum_{t=1}^T \sum_{k \in \tilde{K}_t} \sum_{d=t}^{\tau_k} v_{tk\tau_k} r_{tkd} x_{tkd} - \sum_{s=1}^S p_s Q^s(x) \quad (11)$$

subject to

$$\sum_{d=t}^{\tau_k} x_{tkd} \leq 1 \quad \forall t \in \tilde{T}, k \in \tilde{K}_t \quad (12)$$

$$x_{tkd} \in \{0, 1\}, \quad (13)$$

$$\text{For } \forall t \in \tilde{T}, k \in \tilde{K}_t, d \in \{t+1, \dots, \tau_k\},$$

where for all s ,

$$Q^s(x) = \text{Minimize } \sum_{d=1}^T \sum_{i=1}^4 q_{di}^s y_{di} \quad (14)$$

subject to

$$y_{d1} - y_{d2} = V_d^s - \sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} v_{tk\tau_k} x_{tkd} \quad \forall t \in \tilde{T} \quad (15)$$

$$y_{d3} - y_{d4} = W_d^s - \sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} w_{tk\tau_k} x_{tkd} \quad \forall t \in \tilde{T} \quad (16)$$

where all y_{di} are non-negative for $d \in \tilde{T}, i=1,2,3,4$.

Because all random variables in (M2) are discretely distributed, and their joint distribution has a finite number of realizations, (M2) can be rewritten as the following large-scale mixed integer programming model (M3):

$$\text{Maximize } z = \sum_{t=1}^T \sum_{k \in \tilde{K}_t} \sum_{d=t}^{\tau_k} v_{tk\tau_k} r_{tkd} x_{tkd} - \sum_{s=1}^S p_s \sum_{d=1}^T \sum_{i=1}^4 q_{di}^s y_{di} \quad (17)$$

subject to

$$\sum_{d=t}^{\tau_k} x_{tkd} \leq 1 \quad \forall t \in \tilde{T}, k \in \tilde{K}_t, \quad (18)$$

$$y_{d1}^s - y_{d2}^s = V_d^s - \sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} v_{tk\tau_k} x_{tkd}, \quad (19)$$

for $\forall \notin \tilde{T}, j \in \tilde{J}, s \in \tilde{S}$

$$y_{d3}^s - y_{d4}^s = W_d^s - \sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} w_{tk\tau_k} x_{tkd}, \quad (20)$$

for $\forall \notin \tilde{T}, s \in \tilde{S}$

where $x_{tkd} \in \{0,1\}$, for $\forall t \in \tilde{T}, k \in \tilde{K}_t$,
 $d \in \{t + \mathbb{1}, \tau_k\}$ and all y_{di}^s are non-negative for
 $s \in \tilde{S}, d \in \tilde{T}, i = 1, 2, 3, 4$.

A common attitude in solving NP-hard combinatorial optimization problems (see [4]) is to not insist on optimality but dedicate research efforts to designing fast and high quality approximation methods. A greedy algorithm is chosen for solve the problem (M3). Its robust and implicit enumerative character ensures to achieve the optimal solution or a near optimal solution. In cases like this we can sacrifice the guarantee of optimality that is provided by it in favor of getting a reasonable answer quickly.

Let $E_d = \min\{E_d^s : s \in \tilde{S}\}$, $V_d = \min\{V_d^s : s \in \tilde{S}\}$,
 $W_d = \min\{W_d^s : s \in \tilde{S}\}$, $\bar{v}_{tk\tau_k}^d = v_{tk\tau_k} / V_d$,
 $\bar{w}_{tk\tau_k}^d = w_{tk\tau_k} / W_d$. Then, we get a binary integer programming as follows (M4):

$$\text{Maximize } z = \sum_{t=1}^T \sum_{k \in \tilde{K}_t} \sum_{d=t}^{\tau_k} v_{tk\tau_k} r_{tkd} x_{tkd} \quad (21)$$

subject to

$$\sum_{d=t}^{\tau_k} x_{tkd} \leq 1 \quad \forall \in \tilde{T}, k \in \tilde{K}_t \quad (22)$$

$$\sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} \bar{v}_{tk\tau_k}^d x_{tkd} \leq 1 \quad \forall \in d \in \tilde{T} \quad (23)$$

$$\sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} \bar{w}_{tk\tau_k}^d x_{tkd} \leq 1 \quad \forall \in d \in \tilde{T} \quad (24)$$

where $x_{tkd} \in \{0,1\}$, $d \in \{t + \mathbb{1}, \tau_k\}$, $\forall \in \tilde{T}$,
 $k \in \tilde{K}_t$,

Toyoda's heuristic for the Multi-Dimensional Knapsack Problem (MDKP) (see [6]) starts with no items (or all x 's being zero), and adds one item at a time iteratively as long as the solution is feasible. Following his approach, we propose a heuristic algorithm EBHA that provides a near-optimal solution to (M2) by means of concepts such as penalty vector and effective gradient introduced in [6].

The heuristic algorithm for the mixed integer

programming (M3) is presented as follows.

Algorithm EBHA:

Step 1: Initialization.

Step 1.1: Let $z \leftarrow 0$, $\hat{z} \leftarrow 0$, $x_{tkd} \leftarrow 0$, $\hat{x}_{tkd} \leftarrow 0$.

Step 1.2: Let $K_U \leftarrow \emptyset$, where K_U is the set of accepted items.

Step 1.3: Assign all items to $K_D = K - K_U$, where K_D is the set of items not in K_U and $K = \{k : k \in \tilde{K}_t, t \in \tilde{T}\}$.

Step 1.4: Let $A_U^d \leftarrow (0, 0)$, where A_U^d is the total quantity vector of accepted items in period d .

Step 2: Let $K_C \leftarrow \{k : k \in K_D, \exists d \leq \tau_k \text{ s.t.}$

$(\bar{v}_{tk\tau_k}^d, \bar{w}_{tk\tau_k}^d) \leq (1,1) - A_U^{dj}\}$, where K_C is the set of candidate items.

Step 3: Check K_C . If K_C is empty, goto **Step 7**.

Otherwise, proceed to the next step.

Step 4: Let $\bar{K}_C = \{(k, d) : k \in K_C, d \in \tau_k\}$,

$(\bar{v}_{tk\tau_k}^d, \bar{w}_{tk\tau_k}^d) \leq (1,1) - A_U^d\}$.

Step 4.1: If A_U^d is a zero vector, then we set

$$G_{kd} \leftarrow \frac{\sqrt{2} v_{tk\tau_k} r_{tkd}}{\bar{v}_{tk\tau_k}^d + \bar{w}_{tk\tau_k}^d}, \text{ for } (k,d) \in \bar{K}_C$$

Step 4.2: Otherwise, let $K_U^d = \{k' : k' \in K_U, x_{tk'd} = 1\}$, for $(k,d) \in \bar{K}_C$, we set

$$G_{kd} \leftarrow \frac{v_{tk\tau_k} r_{tkd} \sqrt{(\sum_{k' \in K_U^d} \bar{v}_{tk'\tau_k}^d)^2 + (\sum_{k' \in K_U^d} \bar{w}_{tk'\tau_k}^d)^2}}{\bar{v}_{tk\tau_k}^d \sum_{k' \in K_U^d} \bar{v}_{tk'\tau_k}^d + \bar{w}_{tk\tau_k}^d \sum_{k' \in K_U^d} \bar{w}_{tk'\tau_k}^d},$$

Step 5: Find that item k whose effective gradient is the largest in a period, i.e.,

$$G_{kd} = \max\{G_{k'd} : (k',d) \in \bar{K}_C\}.$$

Step 6: Accept k . Let $K_U \leftarrow K_U + \{k\}$,

$$A_U^d \leftarrow A_U^d + (\bar{v}_{tk\tau_k}^d, \bar{w}_{tk\tau_k}^d), z \leftarrow z + v_{tk\tau_k} r_{tkd},$$

$K_D \leftarrow K_D - \{k\}$, $x_{tkd} \leftarrow 1$. Then, goto **Step 2**.

Step 7: Let $z \leftarrow \hat{z}$, $x_{tkd} \leftarrow \hat{x}_{tkd}$,

$$y_{d1}^s \leftarrow V_d^s - \sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} v_{tk\tau_k} x_{tkd}, \forall \notin \tilde{T}, s \in \tilde{S},$$

$$y_{d3}^s \leftarrow W_d^s - \sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} w_{tk\tau_k} x_{tkd}, \forall \notin \tilde{T}, s \in \tilde{S},$$

$$z \leftarrow z + \sum_{s=1}^S p \sum_{d=1}^T (q_{d1}^s y_{d1}^s + q_{d3}^s y_{d3}^s),$$

$$\hat{z} \leftarrow z, \hat{x}_{tkd} \leftarrow x_{tkd},$$

for $t \in \tilde{T}, k \in \tilde{K}_t, d \in \{t + \mathbf{L}, \tau, k\}$.

Step 8: If $K_D = \emptyset$, the procedure terminates. Otherwise, let $z \leftarrow 0, y_{di}^s \leftarrow 0$, for $s \in \tilde{S}, d \in \tilde{T}, j \in \tilde{J}, i=1,2,3,4$, and proceed to the next step.

Step 9: we set

$$G_{kd} \leftarrow \frac{\sqrt{0.125} v_{tk\tau_k} r_{tkd}}{v_{tk\tau_k}^d + w_{tk\tau_k}^d}, \text{ for } k \in K_D.$$

Step 10: Let $G_{kd} \leftarrow \max\{G_{k'd} : k' \in K_D, d' \in \{t + \mathbf{L}, \tau, k\}\}$.

Step 11: Let $K_D \leftarrow K_D - \{k\}, x_{tkd} \leftarrow -1$.

Step 12: For $d \in \tilde{T}, s \in \tilde{S}$, if $V_d^s \geq \sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} v_{tk\tau_k} x_{tkd}$,

then $y_{d1}^s \leftarrow V_d^s - \sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} v_{tk\tau_k} x_{tkd}$,

otherwise, $y_{d2}^s \leftarrow \sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} v_{tk\tau_k} x_{tkd} - V_d^s$;

if $W_d^s \geq \sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} w_{tk\tau_k} x_{tkd}$, then

$$y_{d3}^s \leftarrow W_d^s - \sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} w_{tk\tau_k} x_{tkd},$$

otherwise, $y_{d4}^s \leftarrow \sum_{t=1}^d \sum_{\substack{k \in \tilde{K}_t \\ \tau_k \geq d}} w_{tk\tau_k} x_{tkd} - W_d^s$;

$$z \leftarrow \sum_{t=1}^T \sum_{k \in \tilde{K}_t} \sum_{d=t}^{\tau_k} v_{tk\tau_k} r_{tkd} x_{tkd} - \sum_{s=1}^S p_s \sum_{d=1}^T \sum_{i=1}^4 q_{di}^s y_{di}^s$$

Step 13: if $\hat{z} < z$, then $\hat{z} \leftarrow z, \hat{x}_{tkd} \leftarrow x_{tkd}$.

Otherwise, $x_{tkd} \leftarrow 0$.

Step 14: If $K_D = \emptyset$, the procedure terminates.

Otherwise, let $z \leftarrow 0, y_{di}^s \leftarrow 0$, for $s \in \tilde{S}, d \in \tilde{T}, j \in \tilde{J}, i=1,2,3,4$, and goto **Step 10**.

3. EMPIRICAL RESULTS

In this section, we implement the heuristic algorithm EBHA and compare its solution to optimal solution or LP (relaxation) optimal solution (as the upper bound for optimal solution). The algorithm has been coded in C++ and run under Microsoft Windows Server 2003 Standard Edition using a Server (Intel(R) Xeon(TM) CPU 3.06GHz and 1.0GB of RAM). CPU times were obtained through the C++ function clock(). To conduct our experiments we used randomly generated instances.

For each set of parameters T and K, we generated 10 random small scale instances, for which optimal solutions can be obtained by CPLEX 8.0. We tested heuristic solutions and optimal solutions or LP optimal solutions for all 10 instances, and tabulated the average relative gap and average computation time. Let z_{LP} be the optimum of LP relaxation, z_H be the lower bound by heuristic and z_O be the optimum of the problem. In table 1, the relative gap g_O is defined as

$$(z_O - z_H) / z_O \times 100\% .$$

In table 1 and table 2, the relative gap g_L is defined as

$$(z_{LP} - z_H) / z_{LP} \times 100\% .$$

Table 1 shows the results obtained for a set of small test problems. Test problems 1 have 2 scenarios, 2 periods and 150 items (cargoes); test problem 2 has 3 scenarios, 3 periods, 4 destination ports and 71 items, and so on. For comparison, the optimal solution has been computed using CPLEX 8.0. As can be seen from table 1, the obtained results seem to be encouraging. The gap between the optimal solution and the heuristic solution is small and the computation time is very short. Table 2 shows the results obtained for a set of large scale problems.

From our preliminary computation experiment, we believe that heuristic algorithm would be a very good candidate for solving the problem in time critical or real-time applications such as capacity allocation in e-booking control problem where a near optimal solution is acceptable, and fast computation is more important than guaranteeing optimal value.

4. SUMMARY

Critical to the e-booking control problem is understanding sea freight marketing environment and optimization technique, its impact on the utilization of available capacity and service route planning for liner operator. Since customer behavior and characteristics are highly varying and not known in advance, we develop a stochastic pricing model which tracks customer behavior as well as the arrival process to maximize profit. On the other hand, we have formulated the capacity allocation problem as the two-stage stochastic mixed integer programming model, and presented effective heuristic algorithm which provide fast and near optimal solution. We also presented experimental results to evaluate the algorithm using a wide range of problem instances. The results strongly suggest that the heuristic algorithm is very effective for time critical tactical or operations level decisions, where a near optimal solution is acceptable and fast computation is more important than guaranteeing optimal value.

S	T	K	Number of variables	Number of constraints	Instances tested	Average relative gap (%)	
						g_O	g_L
2	2	126	268	134	10	0.13	0.22
2	3	121	387	133	10	0.21	0.46
3	4	101	452	125	10	0.78	1.93
4	8	43	472	107	10	1.34	2.41
3	5	91	515	121	10	1.11	1.76
3	7	63	525	105	10	1.26	2.22

Table 1. Results for small test problems

S	T	K	Number of variables	Number of constraints	Instances tested	Average relative gap g_L (%)	Average CPU time (sec)	
							Heuristic	LP
3	3	586	1794	604	10	0.73	0.48	393.79
5	6	286	1836	346	10	2.34	0.28	45.08
4	6	473	2934	521	10	1.22	0.86	146.02
3	7	464	3332	506	10	1.34	0.77	147.74
3	4	886	3592	910	10	0.86	1.47	674.78
4	8	764	6240	828	10	1.12	2.33	2084.95

Table 2. Results for large scale problems

ACKNOWLEDGEMENT

This research is supported in part by the Academic Research Fund and the Centre for E-Business of National University Singapore, and the Strategic Research Programme (SRP) sponsored by Agency for Science, Technology and Research (A*STAR) of Singapore.

REFERENCES

- [1] Arnold, B.C., *Pareto Distributions*, Burtonsville, Maryland: International Cooperative Publishing House, 1983.
- [2] Bremaud, P., *Point Processes and Queues*, Springer Verlag, 1981.
- [3] Cheung, R.K., C.-Y. Chen, "A two-stage stochastic network model and solution methods for the dynamic empty container allocation problem", *Thansp. Sci.*, Vol.32, No.2, pp142-162, 1998.
- [4] Garey, M.R., D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, freeman, 1979.
- [5] Lai, K.K., K. Lam, W. K. Chan, "Ship container logistics and allocation", *Journal of the Operational Research Society*, Vol. 46, No. 6, 687-697, 1995.
- [6] Toyoda, Y., "A simplified algorithm for obtaining approximate solution to zero-one programming problems", *Management Science*, Vol. 21, pp1417-1427, 1975.

E-Logistics Comparative Positioning Model: A Multi-National Enterprise Airline Study

John Hamilton

Director of E-Business, James Cook University, Cairns, Australia 4870
John.Hamilton@jcu.edu.au

ABSTRACT

By capturing the hidden value in the multi-national enterprise's (MNE's) supply-chain in areas such as quality, value and innovation; logistics providers can increase the MNE's operating efficiency and enhance its capital utilization, within and enhanced competitive advantage framework. Today, many multinational enterprises (MNEs) use the global communication channels of the internet to strategically move their global activities to more competitive positions. The trend has been to outsource more and more of their non-strategic business functions, with their logistics providers acquiring these new sources of business. These logistics providers are classified as 1st to 4th party logistics providers [1]. A 4th party logistics provider (4PL) offers the complete supply side coordination solutions for the MNE, plus a degree of demand side coordination service [2].

At its peak, the MNE-logistics provider model activates the entire service value chain. Here the 5PL delivers a value chain level of both outsourcing and service. The complete integration of the value chain logistics provider with its vast array of 'added-value skills' and its block of partnered MNEs, yields a unique, innovative, flexible and highly agile partnership, whereby pathways towards 'sustainable' competitive advantage [3] may be developed, and possibly maintained. The MNE, whilst maintaining its final assembly, branding, research, and innovation functions, is increasingly becoming a service orientated, highly 'front-end' (or customer focused) operation.

This paper develops a technique to ascertain the positioning of MNE and logistics provider(s), and where the logistics provider – MNE integration level may be enhanced.

Key Words: logistics, airline industry, service industry, value, logistics provider, competitive advantage, strategy

1. INTRODUCTION

The role of logistics in business has developed over the past fifty years in both scope and strategic importance [4]. Logistics is the management of acquisition, storage, transportation, and delivery of goods throughout the supply chain (businessranks.com). Logistics strategies have influenced customer selection, product design, alliances and many other core business processes[5], and are closely aligned to the business supply side[6]. In the US the outsourcing of IT jobs has resulted in the creation of twice as many jobs as those displaced[7]. The Information Technology Association of America March 2004 report predicts that total savings of global logistics outsourcing policies will deliver total savings of US\$20.9B by 2008 (up from US\$6.7B in 2003, delivering new competitive realities within the business environment.

To be economically productive in the logistics arena the modern MNE should strategically utilize its service supply chain. Leading MNE's are steadily strengthening their 'front-end', downstream, customer-centric and 'innovative research' functions and supply chain structures[2]. They are aligning their key logistics provider(s) and supply chain(s) into a synergistic and globally competitive, strategic solution-set of operations [8]. Their intense, demand chain driven customer focus has yielded new methods of customization and 'one-on-one' 'customerization' [9]. These leading MNE's have

outsourced more and more of their non-strategic business coordination and functions to logistics providers [2],[10]. The logistics provider has been classified as either a 1st party logistics provider (1PL)) such as the mail room; or a 2PL – like the warehousing function; or a 3PL, [8] – like the full transport and storage system; or a 4PL [1] – like the full transport and storage system for several organizations. The logistics classification grouping is based upon the degree of complexity, and the number of business functions under the logistics provider's control, with the 4PL offering the most complex, integrated, cost efficient supply chain solutions.

Recent logistics developments have incorporated the customer driven demand chain into the peak logistics solution. As the required new metrics are unfolding [11], a 5PL progression pyramid [12] is developing with leading MNE's. This progression pyramid is displayed in Figure 1. The 5PL model may also be aligned with the customer relationship model [13] as displayed in Figure 2. The 5PL solution typifies the MNE's strategic vision to target and enhance its valuable, loyal customer relationship.

The 5PL level of outsourcing and service has been conceptualized as a fully activated demand-supply logistics model - termed the 5PL FADS model [6]. Here the supply chain responds to the customer driven demand chain delivering a value chain solution that moves the MNE and its 5PL to a new internationally

competitive position delivering both low cost and high operational efficiency [3]. The MNE is then able to focus on its core business – its final ‘one-on-one’ ‘customerized’ assembly; its innovations, its research; its super-efficiencies, its internal business processes, and/or its individually targeted marketing solutions.

The 5PL model remains the next step in the progression to total logistics integration. The move to increasingly sophisticated, highly-coordinated logistics outsourcing solutions is often driven by the short term focus of senior management personnel, who have short term tenured positions, and who must deliver rapid results for the shareholders. Companies like Dell, eBay, Amazon, Yahoo are moving down this path with great success.

Warren Buffett[14], offers more cautious longer term solutions – and his business improvement and management strategies are equally successful. However, one of Buffett’s recent strategies was to gamble (an unusual strategy for Buffett) against the US dollar and to move large billions of US dollars offshore. This typifies the dynamic nature of executive strategic management today – many avenues are explored, and logistics outsourcing is but one of these avenues for efficiency growth.

FedEx (www.fedex.com) was just another start-up company. Today it is a US\$25B reliable, responsive, relentless, and remarkable delivering the right mix of transportation, information, documentation and supply chain solutions in a wide variety of global customers in 215 countries (and often within 48 hours). FedEx (today a 4PL) has delivered great savings to its corporate partners. Hence for companies that encounter the customerized one-on-one information imperative logistics outsourcing is a means to create efficiencies.

The higher level PL systems deliver lean logistics solutions. Here the superior ability to design and deliver systems to control movement and geographical positioning of raw materials, work in progress and finished inventories at lowest cost[17] delivers a value stream[18]. These lean logistics systems work best for larger businesses with sufficient resources and bargaining powers to deliver such systems[19].

Successful logistics outsourcing can deliver great rewards. For example, in the US telecommunications industry the total logistics costs for companies regarded as ‘best-in-class’ logistics providers equate to 7.3% of revenue (the medium logistics cost for the industry is 14.3% of revenue). For a \$1B company this equates to saving \$70M annually[16].

Over the next five years, it is predicted that the rate of growth in logistics in the US will be between 15% and 20% per year. In 2003 the value of the US logistics industry was US\$910 B (logistics institute). This paper explores ways for a MNE to recognize its current positioning; and to use this knowledge to move to closer to the apex of the logistics outsourcing model - the 5PL solution. It demonstrates how the MNE may tease out its key levers, into a relevant set of measures. It allows the MNE, its key logistics providers, and other service supply chain partners to determine where their solution set is currently positioned. It then suggests the areas in which the MNE should provide additional focus, and thus move to a more competitive and sustainable position.

Recent developments in networked internet linkages, intelligent database interrogation software, alliance partnerships, and outsourcing have moved the leading edge MNE’s into the 4 PL arena (and towards the 5PL area). To measure their relative or comparative positioning tools like the Four Clusters Logistics Model [6] have been developed.

2. THE FOUR CLUSTERS LOGISTICS MODEL

The Four Cluster Logistics Model developed by Gunesh & Hamilton, incorporates four clusters of outsourcing levers. These are presented in Table 1. The model is displayed in Figure 3. Logistics providers partner with MNE’s and develop improved solutions to non-core outsourced MNE business units or functions. A simple

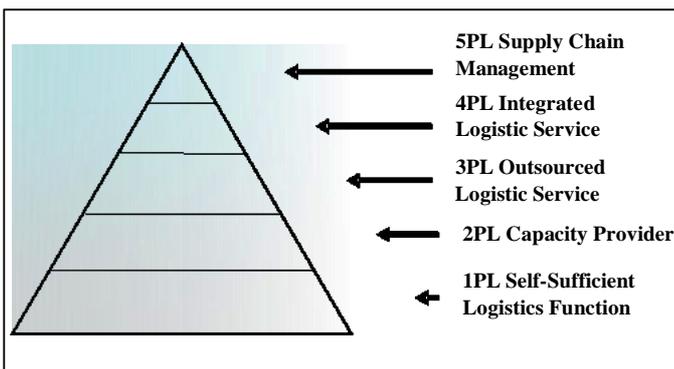


Figure 1: The Morgan Stanley 5PL Model, adapted from Morgan & Stanley[15]

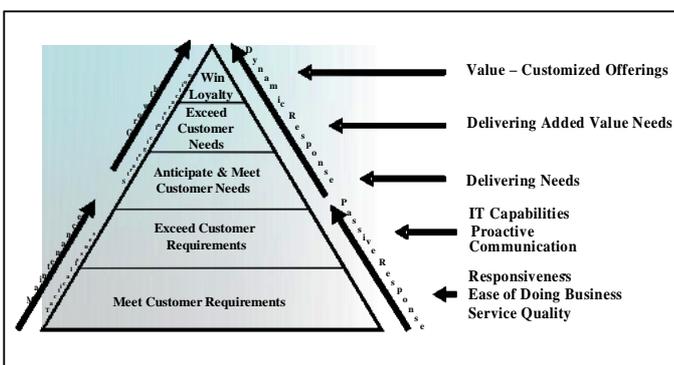


Figure 2: The Customer Relationship Hierarchy Model 5PL Model, adapted from Kuehne & Nagel [16]

The logistics outsourcing solution has been shown to deliver a wide raft of efficiencies, flexibility, cost savings and expertise. For example, thirty years ago

1PL solution may involve outsourcing of the staff canteen, the mail room, cleaning and the like.

the MNE to outsource its logistics functions. A 5PL utilizes all levers, combining and activating them to deliver responsive, value-adding, agile, personalized, competitive solutions to the 5PL-MNE partnership.

Gunesh & Hamilton’s 5PL Logistics Model incorporates these four clusters of outsourcing skills into their current ‘state-of-the-art’ model. They propose that the 5PL opens the value chain pathways for possible ‘sustainable’ competitive advantage.

This model offers the conceptualization of how new value creation methods, lean supply networks, targeted value-adds, additional added value supplier networks, client side connectivity, customer responsiveness, demand driven value chains and the like may provide ways to increase overall profitability. Such a mix of skills provides a greatly reduced scope for direct competition, as it offers complete value chain solutions from the customer to the base level supplier. Some leading edge companies like Dell, IBM and Farmer’s Info are approaching this level of value chain delivery.

3. RESEARCH FOCUS

The Director General of the International Air Transport Association (IATA), Giovanni Bisignani believes the industry must simplify the business [20]. He suggests the rules are outdated and often bilateral in nature, the cost flexibilities and labour structures are outdated, and the lowest cost solution is a non-viable one. He believes operational logistics must become highly automated and interconnected, and that with extensive use of efficient IT (and related technologies) future, flexible, agile solutions.

3.1 Overview

This research uses the Four Clusters Outsourcing Levers. It develops and tests the website interface as a logistics outsourcing tool for the international airlines industry. Questions were addressed such as: ‘Can the four cluster model levers be measured?’ ‘Do these levers provide a way to compare websites?’ ‘Does the four cluster model show the MNE where it may strategically arrange, or change website features to better enable its desired level of logistics?’ ‘What is the relative benchmarked position of the MNE?’ and ‘Should the MNE outsource its value chain logistics?’

To complement the content analysis, a survey and a set of interviews were conducted. The survey assessed consumer perceptions concerning travel industry websites and their service offerings. This survey data was collected from the travel industry and from the travel seeking (or participating) customer perspective. In addition a clutch of interviews helped to provide an understanding of the online travel business requirements from the perspective of their logistics outsourcing to other market players - be they low (1PL) or high level (5PL).

Cluster 1: Technical Levers	Cluster 2: Soft-Skill Levers	Cluster 3: Institutional levers	Cluster 4: Characterization levers
<ul style="list-style-type: none"> Information Technologies Communication Technologies DSS Technologies Organization Technologies IT-Infrastructure Assets Logistics Assets Technical Ability Logistics Knowledge / Know-how SCM Knowledge IT Knowledge 	<ul style="list-style-type: none"> Social Competency Intercultural Competency Interpersonal Competency Communication Skills Organizational Skills Problem-Solving Skills Leadership Skills Team Orientation Empathy Integrity Self-Assurance / Self-Awareness Motivation / Creativity Politeness / Diplomacy Mobility / Flexibility 	<ul style="list-style-type: none"> Government Legislation Industrial Policies Company Laws Financial Services regulation Intellectual Property and Copywriting Policies Competition and Antitrust legislation Corporate Governance and Transparency Transport Policies (Road, sea, air & Rail) Political Stability Economic Growth 	<ul style="list-style-type: none"> Agile Customer Relationship Management Reliable Communication Technologies Integrated Organizational Technologies Database, Web Interface and Knowledge Management Technologies Responsive Database, Web Interface and Activated Business Intelligence On-line ‘Customerization’ and Management Responsive Flexible Demand Chain

Table 1: Four Clusters Logistics Outsourcing Levers, from Gunesh & Hamilton[6].

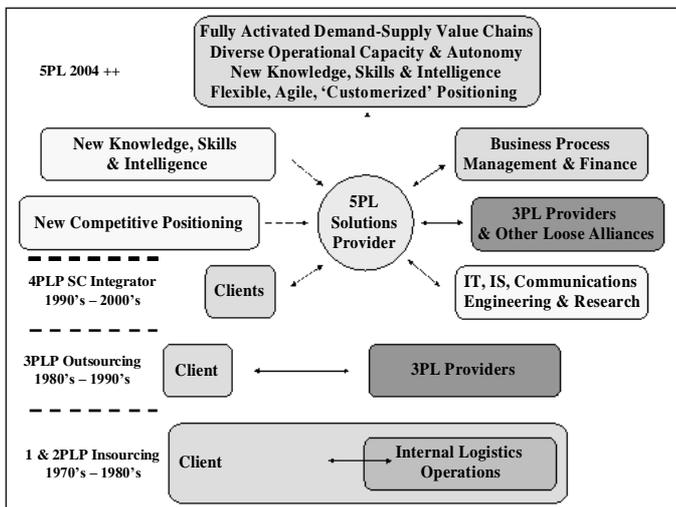


Figure 3: 5PL solutions provider model, adapted from Gunesh & Hamilton[6]

In delivering the various levels of logistics outsourcing, the logistics provider engages an array of levers. The levers displayed in Table 3 constitute ways in which the logistics provider may generate a more competitive position. Measuring the MNE’s ability to activate these levers gives a measure as to the degree of logistics capability the MNE possesses. It is generally cheaper for

Drivers to activate the lever drivers in the four clusters model were identified, refined and added to the model. Multiple data capture methods included: websites and other electronic communications like email, FAQ's; company interviews (marketing and IT executives); and web based surveys of both potential and current customers were incorporated. These additional approaches added greater homogeneity [16] to the study. Mingers[21] and others also suggested that a multi-methodological approach, combining different research methods, yielded greater reliability, and richer research results. The developed drivers were coded, and used to measure the MNE's logistics potential.

3.2 Content Study

Content analysis was used to identify, code, refine and categorize primary patterns in the data[22]. A culling process was used, whereby constant comparative analysis refined the research delivering fuller descriptions and greater homogeneity. The coding of cluster levers was monitored and assessed. A focus group checked the coding of data at two levels – the manifest content level (physical presence of the data), and the latent content level (reading between the lines). Berger[23] suggested this approach is a communication media research technique based on quantitative means.

3.3 Data Coding

Based on prior studies[24], standard data-collection [25] and measurement criteria[26],[27] each MNE website lever feature was binary coded (present (1) or absent (0) against the predetermined model. A score of 1 meant the feature (driver) identified were present, whilst a score of 0 indicated absence. To help limit remaining subjectivity rules a range or a list of acceptable requirements were defined. The elements with performance within the accepted range or meeting the requirements were judged as '1', otherwise they were scored as '0'. For example, page load speed was coded based on industry standards. Thus one ('1') indicated within standard and zero ('0') outside the standard. The major advantage of this coding is that it maintained objectivity. Each site was assessed three times in order to minimise or eliminate judgment variance. Microsoft Excel was used for coding. This research was conducted during May, 2004.

Interpretation of the data offered a mechanism whereby competitive advantage adjustments and/or added-value solutions could be assessed and possibly incorporated into future MNE strategies. Using the above criteria, thirty different, IT-skilled, data-collectors rated each website. Their results were averaged. The relative position of the MNE against its competitors was determined. Major travel agencies (airline website logistics providers) were evaluated and also benchmarked.

3.4 The Four Cluster Outsourcing Levers

The four clusters, displayed in Table 1, each encapsulate

several levers, which, in turn consist of different features (drivers). 100 drivers were identified across the four cluster analysis, and 79 significant measurable drivers were utilized. A sample of the lever drivers identified for cluster 2 is displayed in Table 2.

4. STUDY METHODOLOGY

In this research a predefined logistics comparative positioning model was developed. It was adapted from earlier research[6], regarding virtual benchmarking of website design features and the four clusters outsourcing levers. This model incorporated techniques adopted by worldsbestnetwork.com, w3c.org; and others[6],[24],[26], [27]. These content analysis component techniques provided the basis from which the four cluster logistics driver model may be developed. A multiple lever, multiple feature model, capable of comparing, rating and benchmarking selected websites is developed. Data collection used 34 of the 340 global MNE airlines, and 5 of the 50 high-level global travel logistics companies. These thirty nine websites were systematically analysed to refine, and test this new logistics comparative positioning model.

Greaver's[28] twenty logistics outsourcing factors, and selected elements from Whiteley's 14 parameters model, were used to cross reference Gunesh & Hamilton's lever drivers model. All twenty logistics outsourcing factors, and Whiteley's parameters were suitably incorporated into this study. Hence a tailored comparative positioning model was developed. This model consisted of 79 drivers representing 42 levers.

Cluster 2	
Levers	Drivers
Social Competency	About us
Intercultural Competency	Multilingual, Copyright/trademark
Interpersonal Competency	Returns policy/cancellation
Communication Skills	Email/SMS, Phone, Fax, Contact us
Organizational Skills	ISO, Economic value
Problem-Solving Skills	Calculators/totals/order processing, Customerization
Leadership Skills	Management experience, CIO/Bios
Team Orientation	# Employees presented
Empathy	Treat customer as 1st person, Auto response
Integrity	Guarantee, Delivery time guarantee
Self-Assurance / Self-Awareness	ISO, International stds/awards
Motivation / Creativity	Customer service
Politeness / Diplomacy	Language, Multi-cultural language
Mobility / Flexibility	Tailormade product

Table 2: Lever Drivers for Cluster 2 – Soft Skills Levers, adapted from Gunesh & Hamilton, 2003[6]

Within the second soft skills cluster, 'social competency' is portrayed on the website by 'about us'. This incorporates 'treatment of company info'[27] and organization reasons[28]. Interpersonal competency is exemplified by the returns / negotiation / cancellation policy and communication skills. Mobility/flexibility is viewed via tailor-made product capabilities, incorporating functions like customer seat mapping, and superior provider performance[15]. Similar comparisons

Companies	Clusters				Total	Max	Total Rating	Mean Rating
	1	2	3	4				
Air France	47.6	52.2	57.9	56.3	213.9	400.0	53.5	80.7
Air NZ	23.8	30.4	57.9	37.5	149.6	400.0	37.4	56.5
Australian Airlines	28.6	30.4	63.2	37.5	159.7	400.0	39.9	60.3
Cathay Pacific	81.0	78.3	89.5	68.8	317.4	400.0	79.4	119.8
Emirates	52.4	56.5	52.6	62.5	224.0	400.0	56.0	84.6
Garuda	57.1	47.8	63.2	43.8	211.9	400.0	53.0	80.0
JAL	57.1	56.5	68.4	50.0	232.1	400.0	58.0	87.6
Malaysian Airlines	61.9	56.5	78.9	56.3	253.6	400.0	63.4	95.7
Qantas	90.5	78.3	94.7	81.3	344.7	400.0	86.2	130.1
Singapore Airlines	95.2	78.3	94.7	75.0	343.2	400.0	85.8	129.5
United Airlines	81.0	69.6	73.7	75.0	299.2	400.0	74.8	112.9
Virgin Blue	90.5	73.9	94.7	62.5	321.6	400.0	80.4	121.4
Korean Airlines	81.0	69.6	78.9	68.8	298.2	400.0	74.6	112.6
zuji.com	90.5	82.6	73.7	81.3	328.0	400.0	82.0	123.8
trailfinders.com	95.2	91.3	78.9	81.3	346.7	400.0	86.7	130.9
travelcity.com	100.0	100.0	94.7	93.8	388.5	400.0	97.1	146.6
American Airlines	66.7	65.2	73.7	50.0	255.6	400.0	63.9	96.5
Southwest Airlines	66.7	52.2	57.9	31.3	208.0	400.0	52.0	78.5
Northwest Airlines	66.7	56.5	73.7	50.0	246.9	400.0	61.7	93.2
British Airways	61.9	65.2	73.7	56.3	257.1	400.0	64.3	97.0
US Airways	66.7	73.9	63.2	62.5	266.2	400.0	66.6	100.5
Ryanair	76.2	65.2	63.2	43.8	248.3	400.0	62.1	93.7
Lufthansa	76.2	65.2	31.6	31.3	204.2	400.0	51.1	77.1
Go Fly	42.9	60.9	78.9	50.0	232.7	400.0	58.2	87.8
EasyJet	76.2	60.9	78.9	56.3	272.3	400.0	68.1	102.8
Mean					265.0			
SD								60.6

Table 4: Airline & Travel Agents Web Summary Data

The airlines averaged activation of 63% of possible lever drivers. Consequently, the large logistically focused travel agents were delivering better logistics solutions than those of their MNE airlines. It may be asked ‘why do airlines bother with their websites when they could divert customer traffic to airline travel sites?’

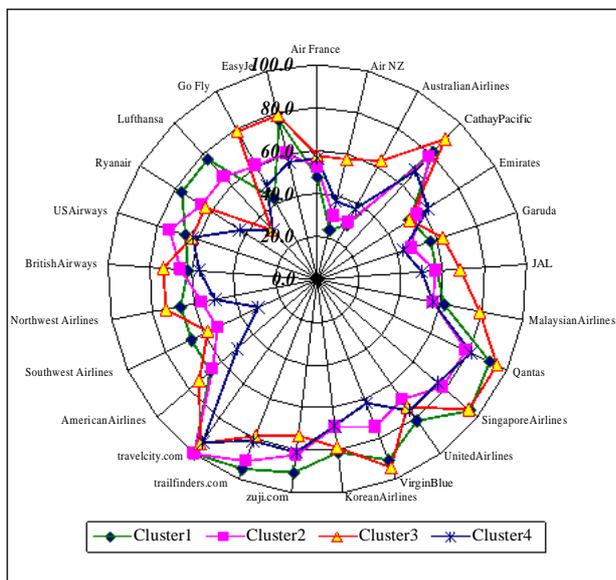


Figure 4: 4 Clusters Logistics Comparison Spiderweb

Figure 4 offers a visual presentation of the relative website logistics performance of various airlines and travel agents to deliver well planned logistics solutions. This study shows that airlines do not plan their websites in a logistically strategic manner. For example, the low level lever drivers are not developed first. There is scant indication of a logistics strategy in most airline websites, and little focus on a transition of relative outsourcing.

Many websites appear to be delivered in an ad-hoc manner. Some airlines do not get the basics right, and others attempt to tackle higher levels of logistics in a poor manner.

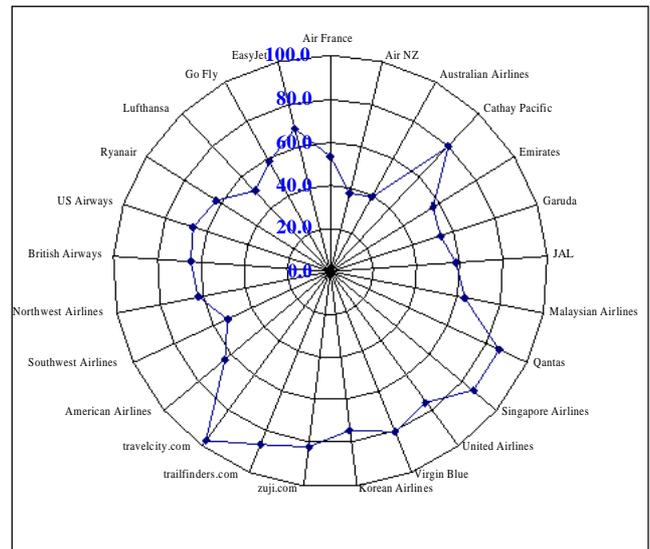


Figure 5: Overall Cluster Rating for selected MNE airlines and airline travel logistics providers

Figure 5 summarizes the logistics presence of the four cluster lever drivers. In this study, only one MNE – Travelocity.com is on-track to progress to a 5PL solution, and two others Zuji.com and Trailfinders.com are operating as 4PL’s. The airlines studied generally rate lower and in the 3PL range, with Qantas, Cathay and Singapore airlines operating in the 4PL area.

6. STRATEGIC INTERPRETATION

Considering the growing global role of e-business, airlines need to consider what they hope to achieve from their web sites. Do they use their site to sell airline seats? If so they should reinvestigate what their core business is, and where are they focusing. From this research, it is apparent that MNE airline travel agent providers offer better solutions via their websites than do the individual airlines. This may be seen as expected, and as a justification for outsourcing! The MNE airline travel agent providers must enhance the customer value proposition, and do so in a cost effective way, in order to compete. They do this by offering added-value products, flight discounts, price and travel comparisons, travel packages, incentives, fast efficient services, and the like. Consequently the airlines have, in the past, used them to save costs and to source more customers - particularly those domestically-based customers, and home country international tourists who intend to travel internationally. Today, the major airline alliances deliver new ways to code share and to strategically move passengers between alliance partners. They have seen their overall passenger numbers grow. However, in the background, the scale and scope of MNE airline travel agents has been increasing. A new problem is emerging – the super MNE

airline travel booking brokerage house (or portal). This super brokerage house strategy is a high risk strategy to the airline industry in the long term. Qantas and other MNE Asian airlines have developed a travel brokerage house (or portal) to help control this situation – ‘Zuji.com’ is their model. This provider offers cheaper fares than can be obtained via the airline’s own website, and also offers comparisons between participating airlines.

This solution could allow Qantas to move all its customer bookings to zuji.com. Qantas website customers would then be diverted to this one central booking site, and a further leveling of the travel booking competition would arise. This may then change the nature of Qantas’s strategic website focus.

The zuji.com website is almost anti-competitive, and appears to be an attempt to block others in the airline travel businesses from capturing this brokerage house domain. It will not stop others, and in time, at least one key logistics brokerage house will emerge.

The zuji.com solution is still small scale when compared to Sabre’s ‘Travelocity.com’ solution. In both cases the airlines are being moved into higher levels of competition, and ultimately the industry will become more low-price driven, with less options for differentiation.

Whilst Qantas, Cathay and Singapore airlines deliver a quality travel service, they do not deliver the same levels of service as those provided by Sabre via its Travelocity.com website.

All large successful travel agent logistics providers are working at above the 4PL level. Hence, for MNE airlines to compete with their own MNE travel agents they must operate at similar levels, or concentrate on improving their supply line efficiencies and outsource these service functions.

The question remains for these MNE airlines where do they choose to focus – on improving their internal logistics, on outsourcing or on improving their core business. This research shows that these large MNE airlines have all selected to move individually and have left gaps in the market. Large scale, globally-connected logistics suppliers have moved in, created a market, and are delivering better solutions than those offered by the individual airlines. Consequently, the airlines are missing an opportunity. To rectify this situation the airlines may choose to:

- 1) reduce the commissions available to travel agents, and control ticketing through their websites
- 2) form a major regional conglomerate and out-compete the logistics suppliers
- 3) concentrate purely on transporting passengers in a cost effective, efficient manner

4) outsource all ticketing and related functions to a large 4-5PL like travelcity.com, thereby relinquishing their sales packages and follow Porter’s[3] 2001 model of cost minimization and efficiency.

This research infers many airlines who are operating without a tight strategic focus, may be more likely to encounter website related passenger ticketing logistics problems. They may become price-driven transporters, and the large MNE 4PL and 5PL logistics-focused airline travel agents could influence their marketplace.

7. CONCLUSIONS

The airline industry operates at a number of virtual levels, and their websites epitomize customer strategy – and none deliver highly customized or ‘one-on-one’ customerized sales-service related solutions.

The airline industry is highly competitive, yet it has not delivered on its prime purpose – to move the maximum number of people it can through its operational spheres of influence. By July 2004 no airline in the world has delivered a quality web-based logistics service to its customers. Each airline maintains a web presence, but this site is often outdated, modified in part at best, and not designed to actually make the passage of winning customers easy. Many sites evaluated, were somewhat user ‘unfriendly’. They were slow to load, some required cookie dumps before one could access them, some do not provide easy customer response avenues, and some did not allow for online bookings. This state of affairs has created a new business opportunity – one where airline travel agent logistics providers can move in, and possibly capture, an additional share of the MNE airline customer sales market. Hence, instead of the MNE airlines controlling, and working with, their chosen logistics providers, the logistics providers may quietly grow in size and develop a small degree of control within the industry.

These airline travel agents have positioned themselves to deliver greater levels of logistics related offerings that those offered by the airlines, hence justification for their existence can be argued as they deliver value added services to the customer, and do so in a more customer-centric manner.

This research indicates that the four cluster model levers can be effectively measured. It shows that airline website drivers can be developed for these levers, and that these may be used to develop a comparative positioning between MNE airlines.

Finally the strategic question still remains - Should the MNE outsource its value chain logistics, and does the size of a business give it protection from being moved towards a lowest price provider?

8. FURTHER STUDIES

On-going research aims to development both a refined computerized logistics model, and a scalable competitive positioning matrix [29].

9. REFERENCES

- [1] Andersson, D., & Jockel, O. "Logistics Competence Provided and Required in Third-Party Logistics Relationships", in Proceedings *NOFOMA* 2002, 569-587.
- [2] Erstad, M. "Commitment to excellence at the ForteHotel Group", *International Journal of Contemporary Hospitality Management*, 2001, 13(7), 347-351.
- [3] Porter, M. "Strategy and the Internet", *Harvard Business Review*, 2001, 79(3), 63-78.
- [4] Rafele, C. 2004, "Logistic Service measurement: a reference framework", *Journal of Manufacturing Technology Management*, 2004, 15(3), 280-290.
- [5] Caplice C. & Sheffi, Y. "A Review and Evaluation of Logistics Performance Measurement Systems", *International Journal of Logistics Management*, 1995, 6(1), 61-74.
- [6] Gunesh, K. & Hamilton, J. "Logistics – A Pathway Towards 'Sustainable' Competitive Advantage for the Multinational Enterprise", on CD ROM Proceedings, ICEB 2003 and in The Third International Conference on Electronic Business (ICEB 2003), NUS, 597-599.
- [7] Greaver 2nd, M. "Strategic Outsourcing", *American Management Association International*, 1999, accessed July 31 2003 at www.amanet.org.
- [8] Voss, H. "International Logistics Outsourcing - Competitive Advantages for 3PL-Providers through Intercultural Competence", in Proceedings, *15th Annual NOFORMA Conference*, 2003.
- [9] Hamilton, J & Selen, W. "The strategic positioning matrix: A path to competitive advantage for on-line, small-to-medium-sized enterprises in the services industry", *Journal of New Business Ideas and Trends*, 2003, 1(1), 20-37.
- [10] Finch, B. "*Operations Now.com*", McGraw-Hill Irwin, 2003.
- [11] Sterne, J. "*Web Metrics*", New York, NY: John Wiley & Sons, 2002.
- [12] Morgan, D. "Focus Groups as Qualitative Research", 2nd Ed., Thousand Oaks, CA: Sage, 1997.
- [13] Kuehne & Nagle "Notrel Networks Leverages 'Miami Gateway to Enhance Customer Service for Latin America", 2003, accessed June 1, 2004 at http://usco.kuehne-nagel.com/cs/KN_case_Nortel_Miami_061504.pdf.
- [14] Cunningham, L. "*The Essays of Warren Buffett*," Revised Ed., NY: Wiley, 2002.
- [15] Morgan Stanley model in Hai, L. & Yirong, S. "An Approach Towards Overall Supply Chain Efficiency - A Future Oriented Solution And Analysis In Inbound Process", Göteborg University, Elanders Novum AB Pub, 2002.
- [16] Anonymous, 2003, "Building Bridges: Relationship Management is the new core competency", Kuehne & Nagel White Paper, Accessed <http://usco.kuehne-nagel.com/>.
- [17] Bowersox, D. Closs, D. & Cooper, M. "*Supply Chain Logistics Management*", NY: McGraw Hill, 2002.
- [18] Jones, D, Hines, P. & Rich, N. 1997, "Lean logistics", *International Journal of Physical Distribution and Logistics Management*, 1997, 27(3/4), 153-173.
- [19] Wu, Y. "Lean manufacturing: a perspective of lean suppliers", *International Journal of Operations and Production Management*, 2003, 23(11) 1349-1376.
- [20] Bisignani, G. 2004, 'Simplifying the Business,' *International Air Transport Association Conference*, Singapore.
- [21] Mingers, J. "Combining IS research methods: Towards a pluralist methodology," *Information Systems Research*, September 2001, 12(3), 240-259.
- [22] Patton, M. "*Qualitative Evaluation and Research Methods*", Newbury Park, CA: Sage, 1990.
- [23] Berger, A.A. "*Media analysis techniques*", Newberry Park, CA: Sage, 1982, 107-110.
- [24] Everett, A. & Paynter, J. "An Analysis of e-Mortgage Sites in New Zealand", on CD ROM Proceedings, ICEB 2003 and in The Third International Conference on Electronic Business (ICEB 2003), NUS, 85-87.
- [25] Sekaran, U. "*Research methods for business: A skill-building approach*", 3rd Ed., Brisbane: John Wiley & Sons, 2001.
- [26] Whiteley, D. "E Commerce Strategy, technologies and applications", London: McGraw-Hill/Irwin, 2000.
- [27] Whiteley, D., Hersey, I., Miller, K. & Quick, P. "Internet e-Commerce: Buying the book and catching the plane", Cape Town, South Africa: *BIT World*, 1999.
- [28] Greaver 2nd, M. "Strategic Outsourcing", *American Management Association International*, 1999, accessed July 31 2003 at www.amanet.org.
- [29] Hamilton, J & Selen, W. "The strategic positioning matrix: A path to competitive advantage for on-line, small-to-medium-sized enterprises in the services industry", *Journal of New Business Ideas and Trends*, 2003, 1(1), 20-37.

E-supply Chain's Implementation in a Transit Economy

Chen Lihua and Yan Tao

Guanghua School of Management, Peking University, Beijing 100871, China
{chenlh,yant}@gsm.pku.edu.cn

ABSTRACT

In the worldwide, firms are increasingly paying their attentions to the integrated Web-based supply chains or e-supply chains because e-supply chains could enhance firms' efficiency and competitiveness. There have been some off-the-shelf products available for supplying activities. But how to establish an efficient e-supply system and how to successfully implement it in the supply process are not as easy as what expected. We propose an implemental and general supply chain framework with various techniques for different supply chain requirements in a transit economy. In a transit economy, the main challenge is to design an e-supply system such that it can be implemented efficiently according to the special supply chain environment of a transit economy.

Keywords: e-supply chains' implementation, efficiency and competitiveness, transit economy

1 Introduction

A supply chain refers to a chain which links each element of the manufacturing and supplying process from raw materials to the end customers, concerning with several organizational boundaries. Supply chain can also be further expanded to include recycling or re-use process. Recently, supply chain has been defined more practically as a connected network with organizations (e.g. suppliers, original equipment manufacturer, distributors, transporters, etc.), resources, and activities of creating and delivering various values in the form of both finished products and services [9] [16] [17] [7]. Nowadays, business firms are more interested in exploiting the integrated supply chains because they want to reduce cost, enhance efficiency and competitiveness, and accelerate the response to the market demand [12][18] [24] [20][26]. As business-to-business transactions based on the internet are increasing, it is becoming critical that firms rely on Web-based supply chains or e-supply chains to provide almost real-time response to market [16] [25] [13]. Thus, many famous IT companies such as i2 Technologies and Manugistics have provided a variety of supply chain software packages and some integrated

suites like mySAP and Oracle li. But for different firms, implementation solution can't keep same. How to figure out this problem? In this paper, we present a practical framework for e-supply chain in a transit economy, which not only can deal with the complexity of operations in e-supply chain system, but also be able to integrate the alliance of the supply chain. In the paper, we consider seven classes of firms with different level information asymmetry. For each level of firms, we proposed a uniform implementation solution based on their available information systems. We use the information asymmetry but not the complexity of operations to class the firms because information asymmetry is more critical than the complexity of operations in e-supply chain management. First, some operations are comparative simple, and need to be managed by IT tools due to the severe information asymmetry. Second, many complex operations are designed to avoid information asymmetry. Finally, eliminating information asymmetry can provide more evaluation to the performance of firms. In a transit economy, firms are developed in a different environment from usual. In many industries there exist some kinds of government intervention, which is

special serious in state-owned firms. So government regulation and role are important factors that shouldn't be ignored in establishing an implemental framework of e-supply chain system. We organize the paper as follows: in the section 2, we first describe the status quo of information system in transit economy which is based on the interviews with many directors of IT companies' marketing departments in a transit economy, and discuss the existed and latent challenges in a highly integrated supply chain from three perspectives: internal challenges, external challenges and government challenges. Section 3 dedicates to propose the implementation framework that identifies various solutions to the different e-supply chain related to the three key factors. Finally, we conclude the discussion and make some remarks about our research.

2 E-supply chain systems

2.1 Characteristic in a transit economy

In a transit economy, usually, many business firms have established considerable foundation in the infrastructure of information systems and in its application, and meanwhile, most business firms have realized that supply chain management can enhance their efficiency and competitiveness. First, hardware and related investment has been almost ready in large or even medium size business firms in a transit economy. Second, information isolated islands are being eliminated and integrated internal operation systems are coming into being. Finally, the concept of supply chain is more and more popular to business firms due to the spreading of media organization and accelerating of the government. However, there still left many deficiencies in a transit economy. First, there are still many firms that do not pay enough attention to IT. Second, few firms have accomplished the strategic enterprise systems. Third, although many industrial chain have coming into being, few firms extend their systems to outside and achieve integration on the entire supply chain. Finally, many firms do not fully realized the importance of e-business and haven't take actual steps to make good use of the internet-based system.

2.1.1 Opportunities

To implement e-supply chain system, there are

some opportunities in a transit economy. First, because industries rapidly develop recently and many industry chains have appeared in a transit economy, it is time to develop e-supply chain systems to enhance the supply chains' as while as firms' efficiency and to provide better customer service. Second, available systems can be easily updated by new technology and off-the-shelf software. Finally, government is going to pay great attention to the industry/supply chain management and to promote the development of powerful supply chain by various supports.

2.1.2 Difficulties

There exist difficulties along with the development. First, the concept of supply chain accepted by most firms is far different from e-supply chain. Second, most firms are not so strong to support the implementation of an integrated e-supply chain system. Finally, there is a lack of leading firms in industry chains that have enough power to integrate the e-supply chain.

2.2 Implementation challenges in a highly integrated supply chain

We illustrate the complexity involved in creating and implementing the integrated e-supply chain system with the two tasks of the system. Both of the tasks were geared towards achieving competitive advantage by reducing inventory, avoiding delays, and providing better customer service. The two tasks are: (1) cooperatively planning among the supply chain and (2) rapidly responding to customer query [18]. Many studies have highlighted numerous problems faced by a firm in adopting complex systems like ERP for internal operations [10]. Similarly, the literature on inter-organizational systems has discussed the problems associated with the implementation of EDI systems that are largely limited to one-to-one relationships [14] [15][19]. But we found that the challenges involved in implementing integrated e-supply chains are more serious than the problems faced by an organization trying to adopt internal systems (like ERP) or inter-organizational systems (like EDI). Integrated e-supply chain systems are the extended enterprise resource planning systems in which diverse organizations need to work together virtually. We classify challenges to the organizations in a highly integrated e-supply chain system into three categories: challenges from inter-

nal of the organizations, those from external of them and those from government.

2.2.1 Internal challenges

Organizations invariably require considerable resources to implement the highly integrated supply chain systems. First of all, the software packages for creating e-supply chain systems are quite expensive and need several powerful computers, database systems, and communications links to support. Secondly, it is particularly expensive to secure the installed highly integrated e-supply chain systems because in such systems: (a) public Internet is used to transport data, (b) many more partners than in traditional EDI systems are usually involved, and (c) data exchanged tends to be highly sensitive. Thus, an organization that implements a highly integrated supply chain system has to expend financial resources that will greatly exceed the expenses on many other information systems. Thirdly, organizations that implement integrated e-supply chain systems need to have highly sophisticated technology infrastructure in place. The IS literature has discussed how organizations get overwhelmed by the technology infrastructure needed for implementing a complex system like ERP [10]. Indeed, the infrastructure that can support a sophisticated and highly integrated e-supply chain system is more technical than others. Finally, the implementation of an integrated e-supply chain system need more reengineering of business process than any other information system does because not only internal business processes need to be integrated, but also these redesigned processes should work synchronously with those of external partners as well. Thus, while the e-supply chain system redesigns the business flows, it must keep accord with its business partners. As given that integrated e-supply chain systems will cause more changes than ERP or EDI, more organizational resistance arises during the implementation of such systems. Senior managers have to involve in the implementing process and devote a lot of time for such system to succeed. Therefore, before a firm creates an integrated e-supply chain, its senior managers have to carefully assess if they have the sufficient commitment, time, resources and abilities to manage the whole process of the firm's reform.

2.2.2 External challenges

Challenges, called external challenges, which are related to bringing supply chain partners on-board and making them a part of the extended enterprise systems are as important as the internal challenges. Some literatures on inter-organizational systems have discussed the importance of business partners in the successful adoption of such systems. For instance, Hart and Grover in [15] and [14] respectively discussed the level of mutual trust which is needed for business partners to be willing to adopt electronic data interchange. In the implementation of an integrated e-supply chain implementation, mutual trust usually gets amplified because of the better relationship among business partners and because of the system's integrating strategies which is a major challenge and can improve firms' competitiveness in the system. While some business partners agreed to become a part of the system for their own benefit, others were reluctant in adopting e-supply chain systems due to resource, technical, relationship (trust and vulnerability) issues, or simply due to a resistance to change to a new way of doing business. On the one hand, the leading firm must have enough power or resources to make the partners adopt e-supply chain systems [15] [14]. On the other hand, in order to implement an integrated e-supply chain system the partners in the system also need large resources and a strong information technology infrastructure, somewhat comparable to that of the leading firm. Further, if the required functionality is built based on the partner's existing information systems, considerable amount of IT expertise for creating the interfaces of diverse systems will be needed. Finding such expertise in some partner organizations proves to be quite difficult, particularly if they are small [15] [14]. Thus, because of the massive financial, managerial, and organizational resources, a firm may give up the aim for a highly integrated e-supply chain. In this way, although the firm can minimize the challenges and reduce the risk of failure encountered during the implementation as while the firm also throws away the opportunities of cooperative development with the integrated supply chain. In order to figure out the paradox, we design some system-based implementation patterns for the different kinds of integrated e-supply chain systems such that the

system can be implemented efficiently with less challenges and risks.

2.2.3 Government Challenges

Above, we have analyzed major internal and external challenges involved in creating a highly integrated e-supply chain system. Now we turn our attention to the government challenges. Notice that many organizations in a transit economy usually have not had the resources, technical infrastructure, commitment, power over their partners, or resourceful chain partners enough to support the implementation of such complex systems, i.e., China is in a lack of powerful leading firms in industry chains. If only relying on the development of firms themselves, e-supply chain systems are too difficult to be implemented, which may delay to improve the firms' competitive advantage and lose the marketing chance. It is crucial for government to focus on some important industries and support them to integrate the whole value chain by e-supply chain system. These industries may include high technologies such as electronics, biotech and spaceflight, strategic industries such as automobile, steel and oil, former state monopolized industries such as banking, mobile and electric power etc. Government should play a different role to support firms at different developing stages to integrate an e-supply chain system. For example, for a newly arisen industry integrating e-supply chain system is unpractical and has no effect. The government should popularize the e-supply chain thought to firms along with favorable policies. When the industry grows up to an extent, there will be increasing need for sharing information and integrating supply chain, but usually few firms have enough power to lead this implementation. The government then can play a key role in promoting the establishment of a suitable e-supply chain system by providing an information platform and issuing some stimulant policies. Finally, how to set up and maintain the trust relationship between government and firms to avoid and punish cheat is a great challenge to a government.

3 An implementation framework of e-supply chain

Traditionally, the issue of e-supply chain implementation was studied based on following two

dimensions: complexity of a firm's internal operations and the firm's ability to integrate their external business partners into the e-supply chain. For complexity of operations, the determination of the suitable level of system's integration is critical [1] [21]. Usually, the complexity of operations is actually caused by information asymmetry. To eliminate information asymmetry rather than only to automate the operations, an e-supply chain system should be launched. Although many complex operations can be designed to avoid some kind of information asymmetry, some other kind of information asymmetry can only be reduced by e-supply chain integration. So, e-supply chain integration becomes a most promising area for the improvement of efficiency to the firm's whole value chain. A firm's ability to integrate its external partners into the supply chain refers to its ability to persuade the partners to participate in the chain, to create necessary infrastructure, and to redesign business processes to match the firm's supply chain system. Two main factors that will facilitate integrating external partners of a firm's supply chain are: trust between the parties and the power over the partners. The ability to persuade business partners to integrate their information systems with leading firm's is considered more important [15] [14]. On the other hand, for a developing country, business firms are not mature generally and they have not enough consciousness and power to integrate e-supply chain. To speed up the integration of e-supply chain in some important industries the government plays a pivotal role. In this paper, we propose the three-dimension implementation framework based on the consideration on information asymmetry, the firm's ability to integrate their external business partners into their e-supply chain and the government's regulation. As discussed above, the first dimension, the information asymmetry can apply to the firm's internal operations, external operations, or both (i.e., it has three levels). For the second dimension, the firm's ability to integrate partners into the supply chain, can be either high or low (i.e., it exists at two levels). For the third dimension, the government regulation in one industry may be strong or weak. Hence, we have $3 \times 2 \times 2$ situations to classify the approaches to implement an e-supply chain. Note that in the situation

where a firm's information are only internally asymmetric, it is not necessary for the firm to integrate external partners and the government regulation will not play any important roles in integrating the supply chain. We, therefore, fold the four cells of our classification corresponding to the internal information asymmetry into one. Additionally, when the firm's ability to integrating is high, there is no need for the government regulation. Thus we fold the two cells of our classification corresponding to the external information asymmetry and high ability to integrate into one. And the same is applied to the two cells corresponding to the internal and external information asymmetry and high ability to integrate. These reduce the number of distinct e-supply chain implementation approaches to seven.

Besides the above issues, managers need to decide whether to build the required e-supply chain functionality based on their existing systems or to buy an off-the-shelf software package to automate and integrate their e-supply chain. Although integrating existing information systems into the new supply chain system can increase the technical complexity of the task, it really can reduce some of the challenges and huge investment which the new system causes, for example, which may need less investment, reduce changes in business processes and changes in procedures, and do not need more trained employees in the use of the updated system, etc. In fact, it is enough for the firms which have already installed certain software to buy additional components of the best-of-the-breed supply chain-related software from vendors, and then customize and integrate them into their existing infrastructure. Therefore, while deciding the approach to create an e-supply chain suitably, it is also very important for a firm to decide whether to build a supply chain system over its existing IT infrastructure or to buy an off-the-shelf software package. We take into account this buy-versus-build option while discussing each of the seven approaches included in our framework. This buy-versus-build decision has been considered to be a major issue in the system development [5][4] [22][23].

3.1 Strategic enterprise systems

A firm that falls in this cell mainly has internal information asymmetry and need more external integration than internal integration. This situation may happen, for example, when a firm has complex manufacturing operations, but most of its components and sub-assemblies are made in-house. Complex manufacturing operations are devised by multiple manufacturing facilities, assembly lines and numerous inventory storages for raw material, work-in-process, and finished goods inventory. How to share information among all these processes? The firm needs an internal integration so as to ensure the smoothly flowing of its information. Since external partners do not play a critical role in the manufacture of the product or its distribution, the firm does not need to concern with integrating business partners with an e-supply chain system. Sometimes, this approach may also be chosen when the firm does not have enough resources to manage both high internal and external integration simultaneously and it believes that it can more easily or effectively handle internal integration. Government here should stand by and give some general guidance. Under this scenario, if the firm's investment in its legacy systems is small or if such systems have outlived their usefulness, the firm may want to buy and adapt a standard software package. A suitable strategy for such a firm is to install strategic enterprise systems such as ERP systems. Since in this situation external linkages are less important, an appropriate e-supply chain strategy for the firm would be to share the output of its ERP system with external partners over a Web-based EDI linkage [2]. Complicated off-the-shelf packages for supply chain planning, execution, and logistics for integration with external partners will not be effective for such a firm. In the above approach of the e-supply chain implementation, the challenges that the firm encounters will be mostly internal, for example, how large amount of resources, how strong IT infrastructure, and how significant change in management effort. The requirement for the system security is limited in such cases because few internal data and networks need to be secured.

3.2 Operational external systems

A firm in this cell has mostly external information asymmetric and it mainly needs external integration while its need for internal integration is limited. This case may happen, for example, when a large number of components and subassemblies for the firm's production are provided by external partners, the firm mainly focuses on in-house assembly. In such a situation, it is important for the firm to create sophisticated linkages with external partners in order to optimize its external value chain. However, if the firm does not have a high ability to integrate business partners into a sophisticated e-supply chain system, it should aim for a limited integration with its partners. Thus, the appropriate strategy for such a firm is likely to create an operational external linkages in the areas like procurement and logistics. Thereafter, the firm can gradually expand the supply chain functionality to more strategic functions like collaborative planning, forecasting, replenishment, and integrated CRM. However, some firms may already have certain valuable legacy procurement and logistics applications that they may be using for achieving integration with partners through traditional EDI. In this case, instead of investing in new off-the-shelf applications and redesigning business processes for procurement and logistics to suit such software, the firm may simply establish a Web interface with these existing applications. During the implementation of this strategy, challenges will be mainly related to external partners because changes in internal operations are minimal. The firm will be required to devote resources to redesign operational external processes with its partners. To secure the supply chain data and network is important because the data are mainly exchanged in the external systems among partners when the firm has highly integrated external linkages.

3.3 Strategic external systems supported by government

Firms in this cell need the same external integration of an e-supply chain as firms in the cell of Operational external systems cell. Usually, the firms in this cell are comparable small in size but large in number. For each firm, it is inefficient to create operational external linkages separately.

But for the entire value chain, a industrial alliance linked by an e-supply chain may lead to great efficiency. Government can play the key role in supporting and organizing the alliance. Another case is that in some industries such as real estate, there will be great revenue due to information asymmetry and firms in these industries are reluctant to implement e-supply chain systems which means the decreasing of revenue. But for planning the whole country's resources, it is urgent to eliminate the kind of information asymmetry. Only the government have the motive and power to push the implementation of this kind of e-supply chain systems. The main options for the government are to establish information exchange platform, to constitute credit systems and to set up information publicizing platform, etc. Information exchange platform is to provide an exchanging channel for operational information. Based on this platform, firms in the value chain can share manufacturing, inventory, logistics and other operational information. Credit system is to constitute trust relationships among firms in a value chain. By this way strategic external systems can be set up in a much shorter period and integrate areas like collaborative planning, forecasting, replenishment, collaborative product design, and integrated CRM, which is called government support strategic external system. Information publicizing platform is mainly focus on the reducing inefficiency in the economy. Therefore, there are mainly two conditions that need the government to intervene. One is to support the setting up of e-supply chain systems to improve the efficiency of the whole value chain in which no firms has enough power to do this in spite of large number of firms in the chain. The other option is to force the implementation of e-supply chain systems to eliminate information asymmetry in some industries with unfair high profit. These two situations may mostly occur in the newly arisen and developing industries and also, there will be a lot of problems and chances in such industries. So the government should pay them enough attention and make them develop in the right direction. E-supply chain integration is a effective way to go through.

3.4 Strategic external systems

Similar to the previous cell, a firm in this cell mainly has external information asymmetry and need external integration with limited internal integration. However, this firm differs from the firms in the previous cell because it is able to more effectively integrate its external partners into the supply chain. The firm falling in this cell should seek supply chain integration with external partners in strategic areas like collaborative planning, forecasting, replenishment, collaborative product design, and integrated CRM. The firm falling in this cell will benefit from strategic external systems like the ones available from suppliers like Dichain, i2, Manugistics, etc. This option is particularly appropriate if the firm has made only marginal investment in legacy supply chain planning software or if its legacy supply chain planning software has run out of its useful life. These supply chain systems can, in turn, be linked with some standard database management software that may be sufficient to handle the firm's simple internal operations like inventory management and accounting. The firm needs to be conservative in its approach toward integration of the internal system, as high internal integration is not really needed in its operations. As such software packages like mySAP and Oracle11i that emphasize high internal and external integration may not be suitable for such firms. However, if the firm already has some valuable installed IT infrastructure, it may follow a different approach to creating its supply chain solution. For example, in the case that the firm may already have a system that it uses for demand planning, in order to create the e-supply chain system, the firm needs to create linkages between the existing demand planner and other internal systems as well as port the demand data over a Web interface for external connectivity, which will be substantial to build the effort on the part of the firm. Local IT solution provider such as Ufsoft, Kingdee, Dichain and Bokesoft are suitable for such firms. Main challenges in implementing systems of this cell's firms will be to get the necessary resources for supply chain integration with external partners in strategic areas like collaborative planning, forecasting, replenishment, and integrated CRM. The firm and its partners will also require technical sophistication to implement and

secure such systems. Additionally, the organization will face the challenge of redesigning business processes jointed with external partners in order to make the implemented systems efficient.

3.5 Strategic enterprise and operational external systems

A firm in this cell has information asymmetry both internally and externally, and, therefore, has a strong need for internal as well as external integration. However, since this firm does not have a high ability to integrate its external partners into the supply chain, it needs to carefully think through its e-supply chain implementation strategy. While it may appear attractive to the firm to create sophisticated e-supply chain systems that integrate well with business partners' systems, in the absence of adequate partner integration, the firm's efforts may not bear much result and may be an unnecessary drain on its resources. An appropriate strategy for such a firm is likely to be to focus on integrating their internal systems and create external linkages in areas like procurement and logistics. It can defer expansion of the supply chain functionality to strategic functions like collaborative planning, forecasting, and replenishment, and integrated CRM. If the firm does not have a strong existing IT infrastructure of value, the firm should consider the option of scrapping the existing systems and building this e-supply chain system from scratch. The firm in that case can create the requisite supply chain functionality by opting for some off-the shelf strategic enterprise software that enables integration of internal processes, and by creating operational linkages for procurement and logistics. On the other hand, if the existing information systems of the firm are adequately serving its needs for internal integration and the firm also has some applications in place for procurement and logistics, the systems building option for the firm could be to integrate procurement and logistics operations with the application for internal operations and establish linkages over a web interface with external partners. Major challenges in implementing systems corresponding with this cell will be related to implementing a strategic enterprise system and integrating its existing legacy systems with it. Thus, it will need considerable amount of resources, technical expertise, and internal pro-

cess redesign. The firm will also be required to devote resources to redesigning operational external processes jointly with its partners and to securing data and the network. However, considering the fact that the data exchanged is not likely to be so strategic in nature, a high degree of trust with the business partners is not very critical for implementing the external linkage.

3.6 Government support strategic extended enterprise systems

The government's role is similar to that described in Government support strategic external systems.

3.7 Strategic extended enterprise systems

Similar to the analysis in the previous systems, the firm in this case has information asymmetry both internally and externally. Therefore, it needs an e-supply chain solution to incorporate both internal and external integration. Moreover, the firm in this cell has high ability to integrate business partners into a sophisticated e-supply chain system. The firm can take two ways to integrate its internal operations and supply chain system: either through an off-the-shelf integrated software package or through a customized solution that integrates best-of-the-breed software packages and ports its output over a Web interface, as done by the electronics manufacturer. Two major initiatives in the off-the-shelf integrated category come from SAP (called mySAP) and Oracle (called Oracle11i). Both mySAP and Oracle 11i are relatively new systems that have not been widely installed. In fact, problems of compatibility with some existing applications that Oracle11i is reported to have, seem to have made some companies hesitant in implementing this software. Thus, the firm's IT experts may have to get deeply involved in the implementation process of these complex software packages and be ready to trouble-shoot some unforeseen problems. On the positive side, both SAP and Oracle have been in the market with their enterprise integration software for a long time and have the necessary expertise to facilitate the successful implementation of their systems. The second scenario might be that the firm has invested already in

some type of systems, e.g., internal connectivity software like ERP or supply chain software from vendors like i2 and Manugistics. Thus, it may not wish to scrap its investment in existing software solutions and start afresh with a solution like mySAP or Oracle 11i. In fact, it might retain of the best-of-the breed functionality offered by software from vendors like SAP, i2, Ufsoft, Kingdee, Dichain, Bokesoft, Manugistics and Siebel instead of committing itself to a single vendor software like mySAP or Oracle 11i. In such a situation, the firm faces with the task of integrating various systems at its end with those of its supply chain partners through a Web interface. This is indeed a complicated exercise and there are likely to be several implementation difficulties. Major challenges in implementing these supply chain systems will be those of integrating multiple systems within the firm as well as integrating them with business partners' systems. Considerable amount of financial resources and technical expertise will be required to make such integration possible and to secure the data and the network. Besides, the firm and its business partners will also be required to undertake substantial process redesign necessary for supply chain integration in strategic areas like collaborative planning, forecasting, replenishment, and integrated CRM.

4 Remarks and Conclusion

Usually, the e-supply chain system is a comparable new concept to the firms in a transit economy. Though the idea is promising and many off-the-shelf products may have appeared in the market of a transit economy, how to implement E-supply chain system is still a great challenge to the firms. It is sure that there is not a uniform solution which can fit the different requirements and the various legacy system bases. This paper mainly analyzed the different roles of firms and government in the implementation of E-supply chain system and proposed an framework to implement e-supply chain systems in a transit economy.

References

- [1] Allen, B. R. and Boynton, A. C., Information architecture: In search of efficient

- flexibility, *MIS Quarterly*, Vol. 15(4), 435-445(1991).
- [2] Angeles, R., Revisiting the role of internet-EDI in the current electronic commerce scene, *Logistics Information Management*, Vol. 13(1), 45-57 (2000).
- [3] Anonymous, Transportation Dot.coms: Fast out of the gate, *Supply Chain Management Review*, S6-S7(2000).
- [4] Arthur, L. J., Quick and dirty, *Computerworld*, December 14(2000).
- [5] Bakersville, R. L. and Stage, J., Controlling prototype development through risk analysis, *MIS Quarterly*, 20(4),481-504(1996).
- [6] Brock, J. R., Supply chain success, *InfoWorld*, April 20 (2001).
- [7] Collett, S., Supply chain tools for all, *Computerworld*, February 22, Vol.20 (1999).
- [8] Cooper, E., *Business research methods*, NewYork, NY: McGraw-Hill, (1995).
- [9] Copacino, W. C.,*Supply chain management*, Boca Raton, FL: St. Lucie Press(1997).
- [10] Davenport, T. H., *Mission critical-realizing the promise of enterprise systems*, Boston, MA: Harvard Business School Press (2000).
- [11] Eisenhardt, K. M., Building theories from case study research, *Academy of Management Review*, 14/4, 532-551(1989).
- [12] Fisher, M. L.,What is the right supply chain for your product? *Harvard Business Review*, 75(2), 105-116 (1997).
- [13] Glushko, R. J., An XML framework for agent-based e-commerce, *Communications of the ACM*, Vol. 42(3), 106-114 (1999).
- [14] Grover, V., An empirically derived model for the adoption of customer-based inter-organizational systems, *Decision Sciences*, 24(3), 603-649 (1993).
- [15] Hart and Saunders, C., Power and trust: Critical factors in the adoption and use of electronic data interchange, *Organization Science*, January-February , 23-42 (1997).
- [16] Henriott, L. L., Transforming supply chains into e-chains, *Supply Chain Management Review Global Supplement*, 12-18 (1999).
- [17] Hickey, K., Java Chain. *Traffic World*, May 31, Vol. 46(1999).
- [18] Hutt, M. D. and Speh, T. W., *Business marketing management*, NewYork: Dryden Press (2001).
- [19] Iacovou, C. L., Benbasat, I., and Dexter, A. S. Electronic data interchange and small organizations: Adoption and impact of technology, *MIS Quarterly*, 19(4), 465-485 (1995).
- [20] Kalakota, R., and Robinson, M. E., *Business: Roadmap for success*, Reading, MA: Addison-Wesley Information Technology Series (1999).
- [21] King, J., Ariba updates e-procurement software with global support, *Computerworld*, August 29(2000).
- [22] Lacity, M. C. and Hirschheim, R., The information systems outsourcing Bandwagon, *Sloan Management Review*, 35(1), 13-25 (1993).
- [23] Lacity, M. C., et al., The value of selective outsourcing. *Sloan Management Review*, 37(3), 73-86 (1996).
- [24] Magretta, J. F., Global, and entrepreneurial: Supply chain management, hong kong style. *Harvard Business Review*, 76(5), 103-114 (1998).
- [25] Mecker, S. S., Internet supply chain management, *Electronic News*, August 16, Vol. 48 (1999).
- [26] Somendra Panta, Rajesh Sethia and Madan Bhandari, Making sense of the e-supply chain landscape: an implementation framework, *International Journal of Information Management*, Vol. 23, 201-221 (2003).

Experimental Study on Effect of Information Behavior by Media Representation Patterns of E-Commerce System

Jiangyu Li

School of Economic & Business Administration, Beijing Normal University, Beijing 100875, China

ABSTRACT

The article made study on effect of different media representation pattern of the hierarchy in shopping-on-line system. The result shows that the effect of virtual reality representation pattern is better than that of Text in higher levels of the hierarchy, but more levels in virtual reality pattern increases the errors of user's activity. It also shows that the media representation pattern of first level have greatest impact on improving performance efficiency.

Keywords: shopping-on-line, virtual reality, hierarchy, media representation pattern

1. INTRODUCTION

In recent years, the development of Internet was growing enormously. While changing working pattern of humans, it also brings in another business model, say online shopping, its unique charming made itself a huge commercial future and even will change current existing traditional way of business model^[6]. The study^[2] found that visual reality pattern of commodity information involved in online shopping system can make buyers to give positive comments to commodity information and effectively arouse the buying desire of consumers. Therefore, that, factors in visual reality of commercial websites affect information behavior and shopping behavior of humans, gains a lot of attention from academic professionals. Those attentions can be interpreted as the following: the proceeding structure of consumers' visual experiences, and the relation between direct experiences of consumers and indirect experiences of consumers in visual reality system^[9]; the applicability of design rules of shops in visual reality system compared to that of real life shops, and the study on the advantages of shops' individuation in visual reality system to consumers^[10]. On the respect of media representative pattern, a primary finding is that representative pattern of shopping system in visual reality is in the better position for consumers to feel familiar in shopping process and have high frequency rate to complete buying tasks compared to that in Text pattern, while on the aspect of representative in systematical structure^[11], how does representative in visual reality affect consumers' information behavior still remains unclear.

As for the study on hypertext model, it was found that the overview was very important for the development of task strategy and working efficiency of the users because it constructed the mental map of users for the system^[3,8]. Representative pattern will greatly affect the development of the overview, even affect the interactive way with the system. One effective method was to fully use established experience and inherent cognitive ability of users, using their familiar objects and activity model to metaphorically demonstrate interface information and

interactive model, which is called metaphor Model^[1]. This understanding and practices had become the major pattern of representative and interaction of current information system. Visualization theory concerned perceived that users' inherent capability of understanding space and space's relationship should be utilized efficiently to allocate information. Therefore, different visualization utilization models are formulated [5,7].

The hypothesis is that hierarchies of shopping system in VR pattern will be easier for users to set up an overview of the system and formulate an effective interactive model, and it will be easier for users to formulate information cognition, but its efficiency of information activities (moves or actions in information space) maybe low due to its high frequency of going around to search objects compared to that in Text pattern, in this sense I suppose to draw a conclusion that VR pattern is better than that in Text pattern in higher level of hierarchy structure; VR pattern with more levels is easier for users to complete the task, but performance efficiency in Text pattern is higher than that in VR pattern, the understanding on the previous conclusion in online shopping environment needs to further test and examine.

2 METHODS

2.1 Subjects

30 subjects are from Beijing Forest University, they are all inexperienced computer and web users with average age of 20.5 years old, 22 of them are male subjects.

2.2 Experimental Devices

PC (Pentium 166), 14' Monitor, two-key Mouse.

2.3 Experimental Materials

To set up three on-line shopping centers The shopping centers use same 4+4+3+1 structures with four levels referring to shopping streets, shops, shopping berth and

commodities respectively and each level consists of four, four, three and one object respectively, say shopping streets have four shops with four shopping berths and each berth consists three items of commodities, and commodities has a list to describe its characteristics and indicate its name. The structure and objects of three shops are all the same. The first level at first shopping center will be presented with visual reality, and other levels will be presented with Text representative (hereinafter referred to as VR100) ; the first and second level at second shopping center is presented with visual reality, and other levels are presented with Text representative (hereinafter referred to as VR11) ; second level at third shopping center is presented with visual reality, and other level is presented with Text representative (hereinafter referred to as VR010) . Entering the first three levels of the shopping center hierarchy will be proceeded by moving rolling bar at the bottom of the screen (for text pattern) or walking (for VR pattern), and then view all objects or the whole shopping site. Shoppers can enter next level of shopping site by clicking object-text, also they can back up the previous level by clicking Exit or Back button. The last levels of three patterns are all the same with text representative. All materials and the facilitation of the experiment are based on the web, VRML, HTML technology.

2.4 Experimental Task

One target object will appear randomly from 48 objects, subjects should go round and view objects in shopping centers by moving mouse. Subjects choose “determine to buy” when they find target object in the fourth hierarchy, which indicates the completion of one object-searching task. Subjects will execute 20 searching tasks as fast and accurate as possible. Reaction time, objects selected and accuracy in different levels made by subjects will be recorded by computer in the whole process.

2.5 Experimental Design

One factor between-subject design was adopted.

Independent variable is referred to different representative patterns of shops, or each level of shops in visual reality, that is to say, which level is presented by visual reality. Independent variable has three levels with three representatives mentioned in experimental materials, each level is matching three treatments respectively. Thirty subjects are grouped into three teams (each team consists of 10 persons) randomly and allocated into three treatments.

Dependent variables are the following

1.Reaction time (RT) : Time subjects spent to complete one object-searching task, totally 20 object-searching tasks in the experiment.

2.Times of mouse movement (TMS) : Actual number of clicking when subjects process the task.

3.Mouse movement error (MME) : It means that the difference between actual times of mouse movement and necessary times of mouse movement for one task. The shopping center was divided into four hierarchies, and subjects must take at least four times of choices to find the object, in this sense, the actual times of one task would be four. Mouse movement error would be the difference between times of mouse movement and four. Controlled variables: Time delay during changes of different hierarchy level of the shopping center, and it takes 9 seconds to appear next level in each representative.

2.6 Experimental Procedures

30 subjects are grouped randomly into three teams with 10 numbers for each team and each team will conduct three treatments respectively (shopping center with three representative patterns). Under each treatment of Text, TCG or VR shopping system, subjects will read wording instruction first, and will perform for an exercise after understanding the instruction perceived by themselves in order to understand the whole tasks and requirement. The content of practices is irrelevant with that of the experimental tasks. Subjects should fulfill ten object-searching tasks during the practices and enter the actual performance of the experiment: as soon as target objects appeared, subjects will begin performing by clicking “Start”, then the first level of shop system, say shopping street, will be showing. The subjects will perform tasks until he or she find the target object and determine to purchase by clicking “determine to buy”, then the primary task is fulfilled, the next target objects appeared, and the second task begun. This process of objects-searching task will be performed twenty times by each subject. The computer will monitor the whole process and record all activities of the subjects.

3 RESULTS

The results are shown in table 1, effective subjects in three treatments are 11 persons, 8 persons and 8 persons respectively.

Review all data first (explore, SPSS 8.0 for Windows) and test the normality of the data, then conduct ANOVA analysis (GLM General Factorial in SPSS for Windows 8.0) . Reaction time doesn't include time delay of 9 seconds, unless it marked clearly.

3.1 Reaction time under different three media representative patterns (RT)

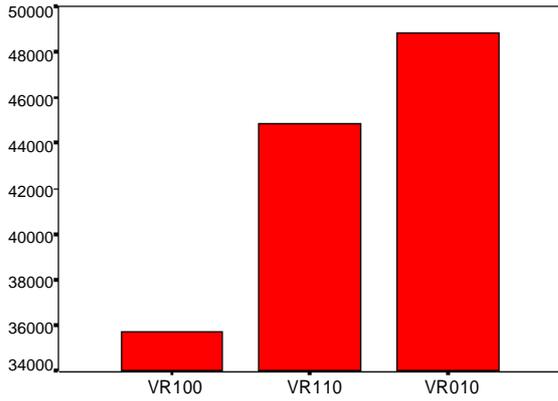


Figure 1. RT under different representative patterns

Table 1 : Reaction time under different representative patterns

	Subject	Mean (ms)	Deviation	Deviation
VR100	11	35685	12367	3530
VR110	8	44896	11141	10565
VR010	8	48905	18944	6313
Total	27	42331	14929	3741

Table1 showed that subjects completed task fastest in VR100, second fastest in VR110 and slowest in VR010.

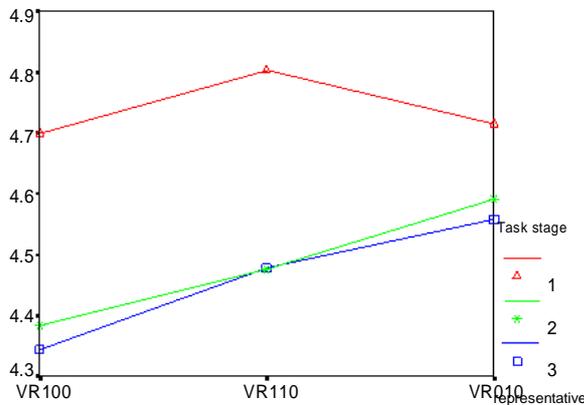


Figure2 : RT of task stages under different representative patterns

ANOVA revealed that there was no significantly difference on task performance among three media representative patterns [F(2,24)=2.161, p=0.137], Post Hoc (LSD) Test found no significant differences among them as well(see figure1).

3.2 Reaction time of each task stage in media representative pattern.

In order to further understand the effect of three media representative patterns on information activity, a deeper research has been made to study the activity of subjects in different stage of tasks in three representative patterns. In this situation, we considered different task stages as additional factor, and divided it into three stages (or levels), first stage with first 7 tasks, second stage with 7

tasks in the middle, and third stage with last 6 tasks. In considering of the requirement of data normality, we conducted changes (RT—>Log (RT+1)). Figure 5 showed that the speed of development under each representative pattern was getting faster as task stages (levels) are developing, which indicated the strong effect caused by practicing. VR110 was the slowest in first stage, and the other two are very close in merit; VR100 was fastest both in second and third stage, VR110 in the middle and VR010 the slowest.

ANOVA indicated that there are significantly differences among three treatments (F(8,72)=5.785,p<0.01) , and the main effect of representative pattern (F(2,78)=4.199,p<0.05) and the main effect of task stages (F(2,78)=0.629,p<0.01) are all significantly different. But there are no significantly differences in interactive function between the two (F(4,76)=0.897,p=0.470) . Post Hoc(LSD) Test found that there are significantly difference between the two of VR110, VR010 and VR100; and significantly difference between the last two stages and the first stage (see figure 2) .

3.3 Times of mouse movement (TMS) 、 Mouse movement error (MME) under different representative patterns

TMS and MME in VR110 are the highest. ANOVA

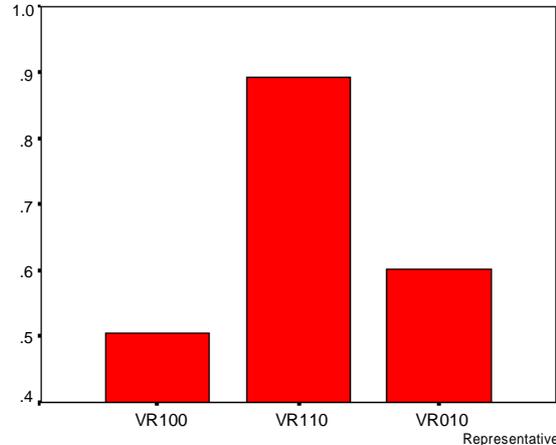


Figure 3. MME under different representative patterns

showed that there was no significantly difference on TMS (F(2,27)=2.828,p=0.077) , and there was significantly difference on MME among the three (F(2,27)=3.914,p<0.05) ,Post Hoc(LSD)Test found significant differences between TMS and MME in VR100 VR 110 (see figure 3) .

4 ANALYSIS AND DISCUSSION

This study showed that using VR in higher level under different media representative patterns results in better performance and higher behavior efficiency of consumers. At the first level, it reaches best performance in VR and worst in Text pattern. This leads a conclusion

that the position of VR appeared is very important under hierarchy structure of three representative patterns. The appearing of VR at first level and second level shows the opposite effect. Using VR at higher levels of hierarchy structure is helpful to lead performance efficiency, but using more VRs at more levels can't lead better performance, but increasing MMS that decreases the speed of the whole task. According to analysis mentioned before, VR at the first level help subjects set up an overview of the system, or setting up a relation between the existing experiences of the subjects and information target and behavior model of the subjects, that is better than that under Text pattern.

Conclusion can be made that VR has a better position to help subjects set up a systematic overview of media representative pattern; the first level in hierarchy structure is more powerful than others to set up a systematic overview of hierarchy levels.

REFERENCES

- [1] Alan Dix, Janet F., Gregory Abowd, Russel Bele, 1993, Human-computer interaction , Prentice Hall, Inc
- [2] Alice Richmond 1996 , Enticing online shoppers to buy——A human behavior study. Fifth international World Wide Web conference, May 6-10, 1996, Paris, France. http://www.w3.org/confrences/www5/fich_html/papers_sessions.html
- [3] Ben Shneiderman 1997 , Designing Information-Abundant Websites: Issues and Recommendations, International Journal of Human-Computer Studies, Int. J. Human-Computer Studies, 47, (1)
- [4] Donna L. Hoffman , Thomas P. Novak , and Patrali Chatterjee, (1995), Commercial Scenarios for the Web: Opportunities and Challenges, <http://209.130.1.169/jcmc/vol1/issue3/hoffman.html>.
- [5] Gary Marchionini 1995, Information Seeking in Electronic Environments, Cambridge University Press, UK.
- [6] GVU,1997, GVU's WWW User Survey, http://www.guv.gatech.edu/user_surveys/servey-1997-04
- [7] Lisa Tweedie, 1997, Characterizing Interactive Externalizations , <http://www.acm.org/sigchi/chi97/>
- [8] P.A. SMITH ,1997,Virtual hierarchies and virtual networks: some lessons from hypermedia usability research applied to the World Wide Web, International Journal of Human-Computer Studies , Int. J. Human-Computer Studies, 47, (1)
- [9] Hairong Li, Terry Daugherty, Frank Biocca, Characteristics of Virtual Experience in Electronic Commerce: A Protocol Analysis, <http://elab.vanderbilt.edu/research/papers/pdf/manuscripts/jim.pdf>, 2001
- [10] Luca Chittaro, Roberto Ranon, New Directions for the Design of Virtual Reality Interfaces to E-Commerce Sites, <http://hcilab.uniud.it/publications/5.pdf>, 2002
- [11] Selected Proceedings, the Second China and U.S. Advanced Workshop in Electronic Commerce 2004, South-west University of Financial & Economic Press

Generic Supply Chain Management System

Man Hing Yu¹, Chee Chern Lim¹, Jesse S. Jin¹, Vera Chung²

¹ School of Design, Communication and I.T., University of Newcastle, NSW 2308, Australia

² School of Information Technologies, F09, University of Sydney, NSW 2006, Australia
{zenoy, chern, jesse, vchung}@it.usyd.edu.au

ABSTRACT

Supply chain management refers to all the management functions related to the flow of materials from the company's direct suppliers to its direct customers. In this paper, we will propose a generic supply chain management system, and describe how the system works in terms of information exchange, workflow coordination and flexible logistic route. We will introduce two data models, which are called PDM and WfMS, and explain how to classify the proposed system into them. And then, we will further describe how the back-end three-layered architecture stores the dynamic data type into the database.

Keywords: Logistic, Supply Chain Management, SCM, E-Business, Database

1. INTRODUCTION

This paper proposed a layered-architecture of database in supply chain management applications. Over the past decade, the rapid advancement in information technology, especially the Internet, has led to the impact of many industries. The Internet environment offers the large flexibility, paperless and 24 hours service to users. As the world-wide-web becomes a popular medium in the human life, people are familiar with the web-based application interface without long time training. While the Internet delivers the large power of coverage, it also provides web tools with the multi-media features.

Supply chain management refers to all the management functions related to the flow of materials from the company's direct suppliers to its direct customers [1,2]. These include ordering, production planning, warehousing, inventory control, packaging and delivery. For example, manufacturing garments will start from purchasing materials from the material supplier, the factory will plan the production material they need. The garments are then shipped to the distributor and finally reach the retailers [3]. The supply chain management involves the collaboration of all members in the entire process.

There are three main scopes that have been considered in our system: information exchange, workflow coordination and flexible logistic routing [4]. The information exchange refers to the information sharing among the supply chain members. The information includes any type of data that could affect other members, such as inventory status, product details and shipping schedules. The workflow coordination refers to the streamlined and automated workflow activities between supply chain members. For example, the manufacture workflow steps involve multiple companies. The time schedule between each step can be tightly coupled, so that collaboration in terms of time, cost and accuracy can be achieved [5]. The flexible

logistic routing refers to find an adaptable and extensible route of the logistic flow of new products. It allows re-structuring the supply chain, such as roles and responsibilities of members, to improve the overall performance. For example, it can be achieved by joining a new partner, inventing a new route of delivery with lower cost or shorter business cycle. These requirements consequently seek for a very generic database schema design, which permits any kind of product data, and logistic information can be stored in the system.

In this paper, we will present our generic supply chain management system, and describe how the system works in terms of information exchange, workflow coordination and flexible logistic route. In Section 2, we introduce the two proposed data models and explain how to classify the proposed system into them. In Section 3, we will describe the design of three-layered database architecture. We will conclude and discuss some current limitations of our approach in Section 4.

2. DATA MODEL CLASSIFICATION

In our proposed system, we classify two kind of data model in the supply chain management: Product Data Management (PDM), and Workflow Management System (WfMS) [6]. This two data model can provide the best solution of the information exchange and workflow coordination in the supply chain management. The PDM is used to control access to documented versions of product designs, which include the traditional single data record such as product details, company contact details. It plays an important role of storage and access of the data and documents throughout the process. Conversely WfMS allow managers to coordinate and schedule the activities of supply chain to optimise the flow of information between the resources and the partners. It is used to coordinate the more complex and repeatable work processes of production.

2.1 PDM (Product Data Management)

Based on this two data model, the system will prepare a data input and retrieval interface for the access of the logistic information online. Since the PDM is an ideal model to control product and contact information, which involve two kind of database structure design: One entry of table storage and multiple entries of table storage. The one entry of table storage satisfy the table schema which have only one-to-one relation with other tables, and the purpose of this table is only used for storing & accessing the record later. For example, the contact management of the supply chain members, it will normally only have one single table called 'member' to store and access the record. However, the multiple entries of table storage have to handle when one table has been normalized by 3-N rules, resulted in splitting into one-to-many relationship with another table. For example, one invoice will have multiple purchased items details, which establish one-to-many relationship with the entry. Hence the sales management in the retailers or the ordering system from the distributors are belonging to this kind of data input.

2.2 WfMS (Workflow Management System)

Conversely, the WfMS category will be more about the time-related task, which can handle any re-structuring, such as roles and responsibilities of all the partners, in the supply chain logistic route. The system can automatically detect the time conflict between partners in the existing route. When the manager arrange for a workflow, which may require the member A, the system will check for the availability of the member A, and wait for guarantee from member A about the delivery in real time. If it has the resource conflicts, the system will inform the user about the next decision.

3. THREE -LAYERED ARCHITECTURE

As shown in the data model above, the application database is required to be able to change the internal storage structure and the access medium. For example, most of the permanent storage medium of the Internet application are the tradition RDBMs, by using the forth generation language to access the database. Alternatively the XML file is used as the storage medium, and it is because of the temporary characteristic of the data or the security risk to store the information into holder's database. On the other hand, the fast growing business activities may require the changing of the database structure, like having additional attribute for customer record. The traditional alternation of the database structure will require large processing time to update the database schema. Consequently, we are seeking a system database model to allow the changing of the storage medium, and also change of the database structure. Three-layered model is to aim for this goal, and to provide the adaptability, extensibility and reusability of the business database

structure. The database connection layer maintains the connection between the system program and the database. It also provides the user interface for the access and manipulation. Secondly, the database structure layer is a run time object layer. It is responsible to manage the table schema in the database. Finally, the physical layer handles the traditional storage medium such as the RDBM database and the XML files. With the integrated power of the dynamic structure database in the back-end, it can allow the user to define corresponding functions to access the data in the database. Hence the system can be platform independent, back-end database independent and also structure independent. Besides of the flexibility, it can also support the multi-media data files, as the user can retrieve the data by content. In addition to the collected data, it also can provide a visualization of the information in graphical vision. With the help of the data visualization technique, the user can understand the current large set in a second.

4. CONCLUSION

In this paper, we have described our approach to supply chain management system based on the two data model, the PDM and WfMS, and the three layered architecture of the database design. We have successfully introduces a supply chain management system in terms of information exchange, workflow co-ordination and flexible logistic route.

Currently, the proposed system is limited by the access time of the database in the structure layer. It is due to every database operation require access of the database structure before the page is loaded. However, the database structure could be possible to store in the run time object, and which can write back to the database when the object is terminated.

In conclusion, the performance of generic supply chain management is an excellent solution in the supply chain management.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support from ARC Discovery Grant (DP0209541).

REFERENCES

- [1] Saarbrucken. (2000): "No E-Commerce without E-Logistics", IDS Scheer.
- [2] Lee, C. B. (2003): "Demain Chain Optimization: Pitfalls and Key Principles", Evant White Paper Series. Evant Inc.
- [3] Gaither, N. & Frazier, G. (2000): "Supply Chain Management", Operations Management, Chapter 15, South-Western/Thomson Learning.

- [4] Lee, H. L. & Whang, S. (2001): "E-Business and Supply Chain Integration", Stanford Global Supply Chain Management Forum (SGSCMF-W2-2001).
- [5] Lee, H. L. & Martin, C. (2001): "Supply Chain Confidence", Global Trade Management, Vastera, Inc.
- [6] Bazan, A., Fetrella, F., Kovacs, Z., Lecoq, P., Le Goff, G. M., McClatchy, R., Murray, S., Solomonides, T. and Vialle, J. P. (2001): "Integrated Data Management and Enterprise Models". Lecture Notes in Computer Science 2118, Berlin: Springer-Verlag, pp.153-163

Hierarchical Structuring of a Workflow Model in Petri-Net

Xinlei Zhao, Yongsun Choi

Dept. of Systems Management & Engineering, Inje University, Kimhae, Korea
xinleizhao@yahoo.com, yschoi@inje.ac.kr

ABSTRACT

In this paper, we introduce the way of deriving hierarchical structure of a workflow model represented in classical Petri-net, even for the cases with cycles, which allows handling a workflow model efficiently. More specifically, our method identifies any block structures as candidates for the subprocesses and represents them as a single block node in the upper layer of the hierarchical model. The proposed method can make workflow analysis and design more accurate and efficient and further lead to a better design on a collaborate environment.

Keywords: Workflow Model, Hierarchical Structuring, Process Abstraction, Hierarchical Petri-Net

1. INTRODUCTION

Recently, workflow management technology has shown to be one of the driving tools in accelerating process-oriented applications ([8], [11], [14]). There are typically two stages of workflow management, workflow specification and workflow execution. The former defines a workflow model, and the latter generates workflow instances guided by the workflow model [15]. The objective of workflow modeling is to provide high-level specification of processes that are independent of the implementation intricacies of the target workflow management system [16]. The recent surges of e-business process automation efforts in the corporations worldwide places workflow analysis and design as a fundamental task of more importance and asks for a model capable of easy interpretation and various analyses with efficiency [4].

Petri-nets, a directed bipartite graph with two node types called places and transitions, have been utilized as an effective methodology in system modeling ([6], [13]). The Petri-net models provide clear graphical representation and profound expressiveness for modeling concurrent, qualitative and quantitative properties. Moreover, the availability of various analysis utilities with Petri-nets provides a means to verify and validate a system, thus has made Petri-nets widely applied for analyzing systems in many areas, including workflow models ([3], [4], [7]). However, introducing the states of the system in the model, called places, the process model represented in classical Petri-nets almost doubles the complexity of representation, in number of nodes and arcs, than the activity-based direct graph model representation. This makes it hard to interpret the business processes represented in Petri-nets for the human designers and rare to be employed by commercial workflow management systems.

Hierarchical Petri-nets, as a type of high-level Petri-nets, provide a mean to model complex system in a more manageable way. When to model complex processes, a classical Petri-net is structured hierarchically by

introducing subprocesses, using either a top-down or a bottom-up approach. This divide-and-conquer strategy of dividing a complex process into smaller subprocesses allows overcoming the complexity for further analysis. Furthermore, the identification of subprocesses provides the way of reusing previously defined processes and often makes it possible to model a complex process more quickly ([1], [2]). Although hierarchical decomposition of a complex workflow process is a useful step in workflow analysis and design, until now there has been no reported work of its automated support in the literature. When the process model is complex, it is quite perplexing for a human designer to recognize substructures of potential subprocesses [5].

In this paper, we introduce the way of deriving hierarchical structure of a workflow model represented in classical Petri-nets, even for the cases with cycles, which allows handling workflow model efficiently. More specifically, our method identifies any block structures as candidates for the subprocesses and represents them as a single block node in the upper layer of the hierarchical model. In case of a model with cycles, our method partitions the given model into several acyclic subsets according to feedback structures, and utilizes the resulting partitioned structures when to compose a block. This paper is organized as follows. In section 2, we present the preliminary concepts such as a workflow-net and its hierarchical extension. Section 3 describes the steps of hierarchical structuring for the workflow models represented in classical Petri-nets, with an illustrative example. Section 4 concludes the paper.

2. A WORKFLOW-NET AND ITS HIERARCHICAL EXTENSION

2.1 Workflow-nets

Petri-nets, represented by a directed bipartite graph in which nodes are either places or transitions, are widely studied and successfully applied in many discrete-event dynamic systems ([6], [13]). Places, represented by circles, describe states or conditions of the system and

transitions, represented in rectangles, describe the events or transactions. The relationships between them are represented by a set of arcs in either direction.

The strong mathematical foundation of Petri-nets and the availability of a wide range of supporting tools have made them popular including workflow domain ([3], [4], [7]). Petri-net-based workflow models represent business logics by a formal but also graphical language. Workflow procedures are specified using a technique with formal semantics of the classical Petri-net and several extensions (color, time, or hierarchy). Availability of abundant analysis techniques [12] is another driving force for the Petri-net-based workflow models. In general, these methodologies can be used to prove properties, like safety, invariance, deadlock, etc., and to calculate performance measures, like response times, waiting times, occupation rates, etc.

A *workflow-net* is a Petri-net-based representation of a workflow process with some syntactical requirements [2]. A workflow process defined in terms of a Petri-net has a single input place *start* and a single output place *end*. And each transition or place should lie on a directed path from *start* to *end*. In other words, there should be no "loose" nodes. Thanks to this requirement, each node can be reached from the place *start* and the place *end* is always reachable from each node, by following a number of arcs. Of all workflow perspectives, e.g., control-flow, data, organization, task, and operation, the control-flow perspective is the most prominent one because it defines the backbone of the workflow on which other perspectives can be specified [4]. Workflow-nets focused on control-flow perspective in modeling a workflow process definition are utilized in this paper.

2.2 Hierarchical workflow-nets

Although the Petri-net has many features to model concurrent, qualitative and quantitative properties, the strict representation of a complex business process is hard to read and understand for the human designers. When facing more complicated situations, the classical Petri-nets become too large and inaccessible, or it is not possible to model a particular activity. The high-level Petri-net, extended with color, time, or hierarchy, etc., helps for more close representation of the problem situation, specific with the perspective considered. The high-level Petri-net, inherits all the advantages of the classical Petri-net, such as the graphical and precise nature, the firm mathematical foundation, and abundance of analysis methods [1].

The workflow-nets encountered in practice have many tasks with very complex interaction structures. For the designer of such a workflow the complexity is overwhelming and communication with end-users using one huge diagram is difficult [5]. Hierarchical Petri-net, as a type of high-level Petri-net can help to model complex situations in more structured and accessible way

[1]. A complex workflow can be decomposed into smaller subflows until the desired level of detail is reached. In addition, this mechanism can be utilized for the reuse of existing workflows [5]. Although hierarchical decomposition of a complex workflow process is a useful step in workflow analysis and design, until now there has been no reported work of its automated support in the literature. Next, we explain our approach of identifying and abstracting the blocks of the workflow model as potential subprocesses to make the process more simplified and manageable.

3. HIERARCHICAL STRUCTURING OF A WORKFLOW-NET

Hierarchical structuring of a workflow-net by abstraction is used to identify the potential sub processes and to make the original process more manageable. Our workflow abstraction method utilizes the concept of inline blocks. An *inline block* is a subset of nodes and arcs among those nodes that satisfies the blocked transition property [15]. According to the Workflow Management Coalition, the *blocked transition property* states that any *inward transition* to the inline block can only occur to the start node of the block and that any *outward transition* from the inline block can only occur at the end node of the block. An inline block is *reducible* to a *block node* or may be modeled as a sub-process of the original process definition. This helps in managing a large-scale model, including verification of structural conflicts [5], being represented as a hierarchy of simple smaller models. Identifying inline blocks manually from a complex workflow [5] is a difficult task even for an experienced process designer.

3.1 Steps of hierarchical structuring

Fig. 1 shows the block diagram of the suggested method for the hierarchical structuring of a workflow model represented in Petri-net. As shown in Fig. 1, our method is comprised of three main steps: 1) Partition a cyclic workflow-net into a set of acyclic substructures; 2) Compose next available candidate blocks; 3) Abstract any block satisfying blocked transition property into a block node.

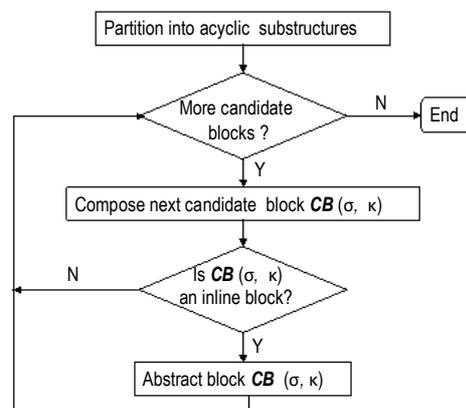


Fig.1. Illustration of the algorithm

Partitioning a cyclic workflow model into acyclic substructures

Step 1 identifies feedback structures and partitions the given cyclic model into a set of acyclic substructures iteratively, even for the models of nested feedback structures. The rank of each node, utilized in Step 2 when to select the next candidate block, is computed with "back edges" [10] temporarily removed. According to the *order of feedback*, nodes are classified into the so-called *nth-order Feedback Nodes*, denoted by *FN*. The *n*-th iteration of Step 1 identifies *FNⁿ*, by classifying the set of nodes *N*. The detailed explanation of Step 1 with illustrative examples is given in [9], with a directed graph representation.

Candidates of inline blocks with cycles

Step 2 first configures potential inline blocks, substructures of the given model, with one of the *Split nodes* (except Feedback Splits) or *Feedback Joins* as the block start node, called the *source*, and one of the *Join nodes* (except Feedback Joins) or *Feedback Splits* as the block end node, called the *sink* [9]. Those potential inline blocks are referred to as *candidate blocks* because they may or may not satisfy the blocked transition property. This initial candidate block, composed with split and join nodes as the border nodes, can be easily extended without further verification effort by adding sequential nodes at the borders. This way of composing candidate blocks can reduce computational cost significantly by focusing on the core candidate blocks.

A candidate block, composed with a node σ as source and another node κ as sink, will be denoted by $CB(\sigma, \kappa)$. For convenience, the set of nodes that spans the candidate block $CB(\sigma, \kappa)$ will be also denoted as $CB(\sigma, \kappa)$, without confusion. The algorithm starts with the simplest candidate block and extends to larger ones, iteratively. More specifically, the algorithm starts with $CB(\sigma, \kappa)$, with σ as one of the candidate sources of maximum rank and κ as one of the candidate sinks, reachable from σ , of minimum rank. At next iteration, new candidate block is selected by fetching new sink κ' of the next higher rank with the same source σ of the current candidate block; when all candidate blocks with σ as source are evaluated or excluded, The algorithm fetches new source σ' of the next lower rank from the stack of candidate sources and proceeds forward.

Blocked Transition Property for Cyclic Workflows

Step 3 checks the *blocked transition property* [15], i.e., no disallowed inward and outward arcs should exist for the newly composed candidate block. If the candidate block $CB(\sigma, \kappa)$ satisfies the blocked transition property, it is abstracted into a new block node, otherwise the proposed method fetches next available candidate block.

In case violations of the blocked transition property happen only at the source or the sink, we can compose an inline block by splitting the source or the sink. Fig. 2 illustrates an example of composing an inline block with $CB(\sigma, \kappa)$ by splitting both the source σ and the sink κ of $CB(\sigma, \kappa)$. Note that if only one of these two violations occur, we can compose an inline block with $CB(\sigma', \kappa)$ by splitting the source σ , $CB(\sigma, \kappa')$ by splitting the sink κ . The arc from node A or the arc to node B, where $A, B \notin CB(\sigma, \kappa)$, does not violate the blocked transition property for the resulting inline block of $CB(\sigma'', \kappa'')$. The newly added nodes σ'' and κ'' in Fig. 2 are null transition of no tasks to perform. The null state nodes σ' and κ' are added to meet the requirements of the Petri-net model, that is nodes should be connected with others of distinct types. Note that it is not necessary the source σ and sink κ are of same node type and is worth for splitting only at the source or at the sink.

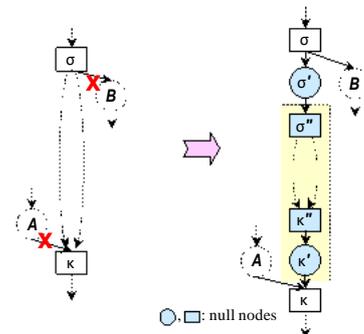


Fig. 2. Composing an inline block by splitting the source or the sink

Types of block structures

According to the node type of the source or the sink, each block structure can be classified into one of the four types that can be reviewed as follows from the Petri-net model perspective:

Transition-to-Transition block (T-T block) or Place-to-Place block (P-P block): The abstracted block node of a subnet, either type of a transition or a state, is connected with distinct type of nodes, and the resulting abstracted model meets the requirement of Petri-net representation. The block structure of any of these types can be represented as a separate subprocess, making the workflow model simpler. When a particular block structure, represented as a subprocess recurs several times in distinct workflow processes, the reuse of this subprocess often makes it possible to model a complex process more quickly. $CB(\sigma'', \kappa'')$ in Fig. 2 is a *T-T block*.

Transition-to-Place block (T-P block) and Place-to-Transition block (P-T block): These types of block structures will make it ambiguous to assess the types of the abstracted block nodes. Moreover, whatever node type is assessed for the abstracted node, the resulting abstracted model, with the abstracted node connected with a node of same type, will not meet the requirements

- [2] Aalst, W.M.P. van der, K. M. van Hee, *Workflow Management: Models, Methods, and Systems*, MIT Press, 2000.
- [3] Aalst, W. M. P. van der, K. van Hee, "Business Process Redesign: A Petri-net-based approach", *Computers in Industry*, 29(1-2), pp. 15-26, 1996.
- [4] Aalst, W. M. P. van der, "The Application of Petri Nets to Workflow Management", *The Journal of Circuits, Systems and Computers*, vol. 8, no. 1, pp. 21-66, 1998.
- [5] Aalst, W. M. P. van der, "Workflow Verification: Finding Control-Flow Errors using Petri-net-based Techniques", In W.M.P. van der Aalst, J. Desel, and A. Oberweis, Editors, *Business Process Management: Models, Techniques, and Empirical Studies*, volume 1806 of *Lecture Notes in Computer Science*, pp. 161-183, Springer-Verlag, Berlin, 2000.
- [6] Aalst, W.M.P. van der. "Putting Petri nets to work in industry" *Computers in Industry*, vol. 25, no. 1, pp. 45-54, 1994.
- [7] Adam, N. R., V. Atluri, and W. Huang. "Modeling and Analysis of Workflows using Petri Nets," *Journal of Intelligent Information Systems*, vol. 10, pp. 131-158, 1998.
- [8] Basu, A. and A. Kumar, "Research commentary: Workflow management issues in e-Business," *Information Systems Research*, vol. 13, no. 1, pp. 1-14, 2002.
- [9] Choi, Y. and J. L. Zhao, "Partitioning into acyclic flows and 2-phased verification of structural conflicts for a cyclic workflow graph", *The Fourth International Conference on Electronic Business (ICEB2004)*, Dec. 2004, *forthcoming*.
- [10] Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms* (2nd Ed.), MIT Press, 2001.
- [11] Delphi Group, *BPM2002: Market Milestone Report*, available at <http://www.delphigroup.com/>.
- [12] Georgakopoulos, D., M. Hornick, and A. Sheth, "An overview of workflow management: from process modeling to workflow automation infrastructure", *Distributed and Parallel Databases*, vol. 3, pp.119-153, 1995.
- [13] Murata, T., "Petri nets: Properties, analysis, and applications", *Proceedings of the IEEE*, vol. 77, no. 4, pp. 541-580, 1989.
- [14] Sheth, A. P., W. M. P. van der Aalst, and I. B. Arpinar, "Processes driving the networked economy," *IEEE Concurrency*, vol. 7, no. 3, pp. 18-31, 1999.
- [15] Workflow Management Coalition, *Glossary. Document Number WfMC-TC-1011*, 1999.
- [16] Workflow Management Coalition, *Interface 1: Process Definition Interchange Process Model. Document Number WfMC TC-1016-P*, 1999.

Manufacturing Enterprises Value Chain Model Analysis within the Context of the Network Economy

Xuejun Xu, Luo Li, Xuelan Hua, Jianhui Huang

Business Administration School, South China University of Technology, Guangzhou 510640, China
Institute for New Industrialization, South China University of Technology, Guangzhou 510640, China
bmxjxu@scut.edu.cn

ABSTRACT

Manufacturing enterprises in the future will have to survive and develop in the network economy with the development of information and network communication technology, the globalization of economy, and the interaction of these two factors. Nowadays the business environments which manufacturing enterprises are facing are very complicated and volatile. The traditional coordinating relationships between operation models of manufacturing enterprises and their business environment have already been changed as a result of globalization, informatization and personalization of customers' needs. New coordinating relationships are in the process of forming.

Concerning operations management in manufacturing enterprise, traditional management theories and methods only confined to the solving of problems within the boundaries of individual enterprise. The introduction of supply chain management makes people considering enterprise problems in a larger context, but most researchers in this field pay more attention to tactic issues. Typical solutions are methods such as postponed manufacturing, information sharing and interests sharing. The systematic and strategic analysis of the supply chain is ignored.

Systematic analysis approach is adopted in this paper. We start our work from the analysis of environment, objective, and structure of supply chain system. Porter's value chain model is then studied and its limitations under new circumstances are pointed out. On the basis of the work described above, we analyze the features of network economy and explore the new value chain model for manufacturing enterprise in the network economy. Finally an empirical case study is conducted.

Keywords: Manufacturing Enterprises, Value Chain, Systematic Analysis, Net Economy, Case Study

1. GENERIC MODEL OF MANUFACTURING ENTERPRISE SUPPLY CHAIN SYSTEM

In this part, we analyze the environment, boundary, function, objective and the structural element of supply chain system.

The environment of system includes behavior of competitors, change of customers' demand, change of policy, trend of technological change, supply condition, etc. Enterprise environment is the part that is usually beyond the control of enterprise.

The boundary of system consists of supply boundaries, distribution boundaries and other boundaries between manufacturing enterprise and its partners. In essence, system boundaries are the boundaries of resources and technologies. The resources and technologies within boundaries are those that can be allocated and managed by manufacturing enterprises.

The function of system refers to the abilities to provide satisfactory products and services to the customers through cooperation and coordination among supply chain partners in response to the requirements of market competitions as well as customers' needs with regard to quality, price, delivery and innovation. The functions of system mainly embody the rapidly responsive ability to

meet the needs of customers with low cost.

The structure of system refers to the relatively stable relationships among subsystems or elements within system boundaries. The relationships here are those that form among manufacturing enterprises and their suppliers, distributors and other partners around the activities to provide products and services to meet the needs of customers. These relationships include not only the relationships with regard to the material flow, information flow and cash flow, but also the relationships concerning personnel and organization, technology cooperation, interests allocation, and performance management among different entities.

A system can survive and develop only by adapting its functions and objectives to the changes of environment. The environment of system determines the objectives of system. The objectives of system raise the requirements of system configuration, while the realization of system objectives depends on the structure of system.

2. VALUE CHAIN MODEL OF MANUFACTURING ENTERPRISE

Michael E. Porter's value chain model reflects the relationships among the environment, objective and structure of the supply chain system of manufacturing

enterprise within the context of traditional industrial economy.

Porter introduced the concept of value chain in his book entitled *competitive advantage*. As a system model, value chain model mainly has the following connotations:

- ò The competitive environment of enterprise consists of five competitive forces including the entry of competitors, the threat of substitutes, the bargaining power of buyers, the bargaining power of suppliers, and the rivalry among the existing players. The five forces determine the achievement of enterprise objectives.
- ò The objectives of enterprise reflect its abilities to gain profit. Enterprise achieves its objectives through value chain configuration and wide variety of value creating activities. Value chain refers to a collection of different activities including product development, production, marketing and sale, delivery and other supporting activities. Value chain and value creation of the chain reflect the relationship between structure and objectives of the system.
- ò Activities along value chain are closely interconnected. Value creating activities of an enterprise can be categorized into two types - primary activities and supporting activities. Primary activities refer to all kinds of activities that associate with material creation of product, marketing and sale, and after-sale service. Primary activities include inbound logistics, operations, outbound logistics, marketing and sales, and service. Supporting activities support primary activities and provide procurement, technology, human resource and other functions within enterprise to ensure mutual support. Supporting activities include procurement, technology development, human resource management, and infrastructure. The former three supporting activities connect to every primary activities and support entire value chain. Enterprise infrastructure also supports the entire value chain though it does not connect to specific primary activity.

Porter's value chain model has the following features:

- ò Partners such as suppliers, customers and distributors are viewed as rivals.
- ò It is considered that the activities along the value chain should be controlled to prevent the invasion of competitors upon market profits. Vertical integration is adopted as primary approach to realize control.
- ò Production activities are managed as basic value generating part, as a result of the enlargement of production scale and the reduction of production

cost are considered as primary management approach.

3. VALUE CHAIN MODEL ANALYSIS WITHIN THE CONTEXT OF NETWORK ECONOMY

The limitations of value chain model of Porter include the following three aspects:

- ò The analysis of environment is simplified. The competition is emphasized while cooperation is ignored.
- ò The recognition about objective is limited. Only the profitability of individual enterprise is taken into account while the identification of customer's needs is ignored, due to the ignorance of the fact that the basis for an enterprise to survive and develop is meeting customers' needs.
- ò The analysis of system structure is limited within enterprise. Dynamic change and interaction, roles of information, innovation and people are ignored.

Therefore, it is necessary to analyze the environment, objective, and structure of the supply chain system of manufacturing enterprises under the new circumstances.

The new environment has the following features:

- ò Globalization. The scopes of competition and resource collocation change from region to the whole world. Division of work and cooperation around the world accelerate the trend of specialization.
- ò Personalization of customers' needs. Mass markets are taken place gradually by segment markets.
- ò The development of information and communication technology infrastructure. Information and communication technology infrastructure provide a platform for inter-enterprise cooperation, making it possible to cut down substantially the management cost within organization and the transactional cost between organizations.

The objective of the existence and the development of supply chain system remains value creating. The value of product is finally determined by customers. It is the measurement of customer satisfaction.

Supply chain is a collection of many enterprises. The factors that can coordinate behavior of different enterprises come from external environment. The most important factor that determines the objective is the needs of final customers. Objective can be denoted as specific output that can meet customers' needs at specific time with specific price and should be comprehensively measured from several aspects including price, time, quality and service.

In order to adapt to the change of environment, manufacturing enterprises should establish new supply chain system structure.

Wide application of information technology is changing the way of communication within and between enterprises, making it possible for enterprises to extend their scope of control and management. At the same time, the trend of specialization around the world has brought about new opportunity for manufacturing enterprises to reconfigure their different kinds of operations activities. For example, procurement, product development and technical service that are previously considered as supporting activities had now become the primary activities that create value for enterprises. Infrastructure that needs lots of investment can be cheaply obtained outside the enterprises.

Under the new circumstances, enterprises can gain profits by establishing tight and diverse relationship with suppliers, customers and other partners. In order to do so, the following two factors are very important. First, enterprises should have their own core competencies, yet it is not necessary to allocate resources on each part of the value chain. Second, enterprises should establish good information infrastructure and international operations platform so as to integrate and make use of resources from all over the world.

The model of relationships between new environment, objective and structure of the supply chain of manufacturing enterprises is shown below.

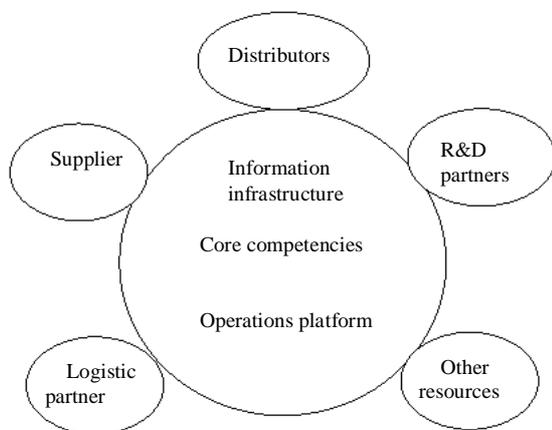


Figure 1: Supply Chain System of Manufacturing Enterprises in the Network Economy

4 A CASE STUDY OF BANKING EQUIPMENT MANUFACTURER

4.1 Background of BEM

BEM (a virtual name) specializes in the development, production, sales and after-sale services of ATM (Automatic Teller Machine) equipment whilst engaging in other monetary electronic equipments. With the initial capital of RMB 50 millions and 12,000 square-metered manufacturing center, BEM can roll out

more than 6,000 sets and generate RMB 100 millions as annual sales revenue. BEM is pinning its long-term hope to become a world ATM manufacturer and it already gains its sound position in China.

BEM owns its state-level enterprise technology center where a handful of veterans work together. Blended with its sophisticated management and trustful alliances with well-renowned suppliers at the home and abroad in ATM field, in addition to its combined strengths in modular design and famous brand components, BEM provides tailor-made products and services of superior quality for different kinds of customers.

Currently, BEM is tapping into the fast-growing economic region Guangzhou and strongly penetrates in Southern China. Besides, BEM further moves into seven Chinese regions by launching 19 contact points for its after-sales services. Due to the intensified domestic competition, BEM relentlessly expands its presence in new-fangled international markets, mainly in Asian developing countries such as Philippines, Sudan and India.

(1) Tremendous Market Potentials

Automatic Teller Machine (ATM) was a novelty to cash-taking China when first introduced in late 1980s. By some statistics, the United States leads the ranking of ATM installations and its deployment exceeds 1,250 terminals per million. Meanwhile, the deployment rates in HK and worldwide are 410 and 204 per million respectively. With a current base of 50,000 ATMs (or 42 sets per million), the Chinese ATM market has just begun to take shape and its all-out explosive growth will possibly make itself the world largest and juiciest by 2007, as suggested by some experts.

(2) Intensive Competition

The hope-fueling China with its untapped potentials has drawn feverish competition amongst a handful of international ATM vendors encompassing NCR Corp., Diebold, Siemens, Fujitsu and Hitachi as well as numerous domestic ones (such as Lide, Guangli and Eastcom) – all vying for a largest piece of pie in Chinese ATM market. Amongst them, the “foreign brands” dominate the market and have grabbed a 90% share, only having slipped the remaining 10% away from their big hands onto their domestic counterparts. BEM controls 70% of domestic brand’s share, making up a 7% of market share in China. Based on the above figures, we still discover that BEM has its dominant position over the domestic brands though its share is obviously small compared with those of international vendors.

(3) Critical Success Factors for Industry

With regard to the unique business environment in China, brand image has become one of the principle factors to ensure the ATM success. In fact, those international super-brands including NCR and Diebold,

as opposed to domestic brands, have been identified as superiority in Chinese state-owned banks and are deployed by most of them based on their brand preferences. Besides, product and service quality are considered to be high performance gaps. Furthermore, Guanxi (i.e. interpersonal relationships) connections underpinned by Chinese cultural practices should be singled out as an important instrument in accounting for the business success in China.

Compared with heavyweights, BEM can provide flexible products at reasonable prices. However, it is still a long coming for BEM to develop strong brand equity and highly reliable products for those banking institutions.

(4) Operating Performance of BEM

After considerable improvement efforts, BEM has achieved significant performance outcomes in 2003, which are illustrated as below:

- (i) **Gross Output Value of Industry** : RMB 200 millions (+63.3%);
- (ii) **Sales Revenue**: RMB 180 millions (+51.6%);
- (iii) **Profit Growth**: Exceed the Directors' annual target.

4.2 Supply Chain/ Value Chain Model in BEM

World-class supply chain configuration has become a crucial issue in the competitive strategies of ATM manufacturers faced with big propositions of cutting their costs and improving operations efficiency in the ATM field. The paradigm shift toward supply chain excellence has promoted the urgent needs of the BEM to relentlessly improve its supply chains for securing its competitive advantages. In order to understand the current situation of the BEM's supply chain management, we will identify how BEM manages its *purchasing, inventory, sales and information*, following the discussion on its supply chain structure.

(1) The Supply Chain Structure

Before a closer investigation on its supply chain in the four perspectives, we will unveil the *supply chain structure* of BEM by demonstrating a simplified web-centric model, which is illustrated as follow Figure 2.

(2) Make-or-Buy

The contemporary economic landscape in the ATM field requires BEM to adopt flexible sourcing strategies in terms of the combinations of in-house production (make) and sub-contracting (buy) in the production and assembly levels. BEM has taken considerable steps towards outsourcing for traditional electronic components (such as display and power) in order to maintain flexibility and avoid over-investment.

On the other hand, BEM assembles the components for the modules, card readers, dispensers, receipt printers, journal printers and diagnostic panels by

itself or sometimes by the sub-contractors located in the Pearl Delta Regions.

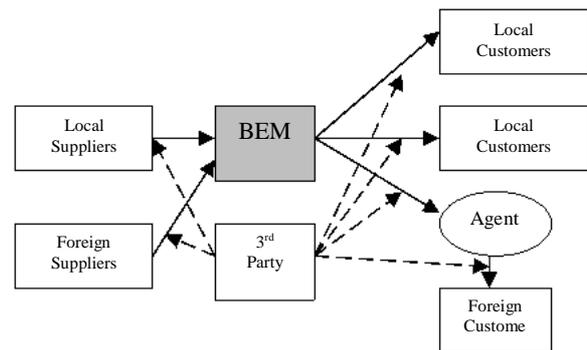


Figure 2: Supply Chain Structure

(3) Supplier Integration

The turbulent Chinese ATM market and ever-demanding banking customers have echoed the needs for *supplier integration*, a managerial approach to tighten the supply chain in order to cut costs and lead times, as well as to build flexibility. Despite the importance of supplier integration, BEM only assigns the suppliers with the responsibility of assembling the generic components while conducting the QC tests by itself.

(4) Inventory

Inventory management is key to optimizing the inbound logistics of BEM, and it is pivotal in the supply chain management. In our section, we put an emphasis on the “inventory structure” and “inventory turnover” of BEM.

First, the *inventory structure* of BEM conforms to the industry standard that requires more raw materials than semi- and final- products. Actually, the raw materials contribute to 40% of total inventory in BEM whereas the semi- and final-products take the account of 30% and 30% of total inventory in BEM respectively. Besides, its lead times are 15 to 30 days, much less than the industry level (i.e. 30- 45 days). Second, the *inventory turnover* of BEM indicates much room of improvement in BEM because it is 72 days, much greater than the industry level (i.e. 40 days).

Having established good relations with its suppliers, BEM has similar VMI arrangements with some of its suppliers. However, BEM faces a vexed issue of how to improve its inventory levels and financial profile given much cash are buried in the inventory in an attempt to shorten the lead time .Figure 3 illustrates the true VMI:

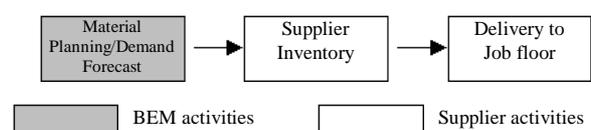


Figure 3: Revised Replenishment Cycle of Raw Materials

(5) Sales

Supply chain begins with the sources of supply and ends with the customer consumption; therefore, fulfilling the customer requirements becomes an integral part of supply chain process to achieve operational efficiency and to produce sales.

BEM has broken into the untapped ATM market in China and aggressively dived into scalable domestic banking institutions encompassing four state-owned commercial banks. Because of good relationship with the Agricultural Bank of China, BEM has obtained the bank's aggregate order that accounted for 22% of its total sales in 2003. In order to support its international strategies and achieve its marketing excellence, BEM has developed the domestic and overseas marketing divisions to manage the corresponding customers respectively.

There has little doubt that what ATM customers really want is excellent after-sales service because technical problems would seriously affect the patronage of bank customers (customers' customers). In order to secure a high level of brand loyalty, BEM has developed 27 maintenance centers based on the product distributions. Besides, BEM provides service guarantees to the bank institutions. According to the service guarantee, BEM will deliver the maintenance service within 4 hours and solve the technical problems within 12 hours.

(6) Information

In facing today's extremely challenging business environment, creating real-time supply chain visibility is becoming very critical for BEM's success in the marketplace. Supply chain visibility refers to the sharing of information to a member within the supply chain, for monitoring, controlling and changing supply chain strategy and operations, from production to delivery (Schoenthaler, 2003). Meanwhile, the challenge of getting the right information to the right person at the right time can be simplified by deploying the appropriate IT applications (Schoenthaler, 2003). Indeed, BEM is only concerned with *the sharing of purchasing and production information with the suppliers* by means of email, telephone and fax. Internet is seldom adopted.

Furthermore, despite the increased expenditure on supply chain integration with the goal of improving customer interaction and reducing inventory costs, BEM

appears lagging behind at the moment given its relatively *low level of information sharing with customers*, resulting in limited collaborative activities such as new product development with customers. Presently, the company relies on fax, phones and emails as major correspondence. It can be understood partly because its customers may perceive impersonal communications a better way to explain their individual needs, and partly because BEM hasn't discovered the tremendous opportunities derived from internet and information technology. BEM should get out of the way of relationships between customers and suppliers while providing a technology-enabled framework for a greater level of integration.

Note: A team from South China University of technology joined the project. We have analyzed the 'AS-IS' of BEM and proposed a 'TO-BE' proposal. The primary solution of the proposal is to reconfigure the previous ERP system and integrate it with customers and suppliers to facilitate process integration and improve information visibility of the supply chain.

ACKNOWLEDGEMENT

This paper is supported by the National Natural Science Foundation of China (No.70272047) and Society Science key Project of Guangdong Province.

REFERENCES

- [1] H. Min, G. Zhou. Supply chain modeling: past, present and future [J]. Computers & Industrial Engineering, No.43 pp231-249 2002
- [2] Ron Ireland, Robert Bruce. Only the beginning of collaboration [J]. Supply Chain Management Review, September/October: 80-88 2000
- [3] D.J. Thomas, P.M. Griffin. Coordinated supply chain management [J]. European Journal of Operational Research. No.94, pp1-15. 1996.
- [4] Schoenthaler R, "Creating real-time supply chain visibility," *Electronic Business*, Vol.8, No.29, 2003
- [5] Disney S.M. and Towill D.R., "The Effect of Vendor Managed Inventory (VMI) dynamics on the Bullwhip Effect in Supply Chains," *International Journal of Production Economics*, Vol.2, No.85, pp 199-215. 2003

On Inventory Strategies of Online Retailers

Frank Y. Chen², S. H. Hum¹, Cheryl H. Sim³

¹ NUS Business School, The National University of Singapore,
1 Business Link, Singapore 117592. E-mail: bizhumsh@nus.edu.sg.

² Chinese University of Hong Kong, Hong Kong.

³ The Singapore National Library, Singapore.

ABSTRACT

This study focuses on inventory strategies of Internet retailers (etailers). The etailer faces options of holding her own inventory or outsourcing through the third party(ies). We assess etailer inventory strategies through mathematical modelling and numerical experiments. When ordering and holding her own stock, the etailer has full control of the order fulfillment process but bears the inventory-related risk. When outsourcing stock, etailer's orders may not get an equal priority as for those of the third party's own. Built upon simple operations research models, the numerical experiments suggest that the etailer is better off relying on others to fulfill orders if her demand (profit margin) is low, but should revert to the strategy of maintaining her own inventory if her sales volume (profit margin) is relatively high. Other factors are also investigated. These findings seem to confirm what are being practised in the industry.

Keywords: Electronic Commerce, E-Retailer, Fulfillment, Inventory.

1 Introduction

This research focuses on online retailers (*etailers*, hereafter) who engage in the sale of physical goods. In general, the order fulfillment process of such an etailer can be depicted by Figure 1.

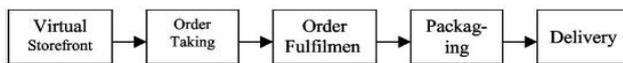


Figure 1 Order Fulfillment Process of Online Retailers

As presented in Figure 1, customers will place an order at the etailer's website. Upon receipt of each order, the etailer will confirm whether or not stock is available to fulfill the order. Assuming that inventory is indeed available, the order will be packaged, after which it will be delivered to the customer. In general, each etailer's order fulfillment strategy comprises three primary components: inventory strategy, packaging, and delivery. It is now well understood that the order fulfillment process for etailers is different from that of traditional bricks-and-mortar retailers (*retailers*, hereafter), and the importance of order fulfillment to etailers is widely recognized (see Ricker and Kalakota, 1999).

By operating in virtual space, etailers are presented with new options in their order fulfillment process in terms of their inventory strategy, packaging, as well as delivery decisions:

Inventory Strategy: Unlike the traditional retailer, the etailer faces several options in terms of her inventory strategy:

- (i) maintaining her own inventory;
- (ii) maintaining zero inventory, and co-operating with one or more third (3rd)-party suppliers. For each customer order the etailer receives, she turns to one of these suppliers to obtain the exact number of units of the goods required to fulfill the order. These 3rd parties can be manufacturers, wholesalers, and other bricks-and-mortar retailers.
- (iii) a hybrid of (i) and (ii). For instance, an etailer may maintain her own inventory as well as rely on 3rd-party suppliers for supplying the product that she sells through her webstore.

Packaging: If the etailer maintains her own inventory, it is most likely that packaging will be performed in-house by the etailer before delivery to customers. However, if she decides not to carry inventory but rely on 3rd-party suppliers to meet customer orders, then there are two alternatives for packaging: (i) the packaging function to be performed by the 3rd-party supplier; or (ii) the etailer continues to perform packaging in-house. In the latter case, after a customer order is received, the etailer turns to one of her 3rd-party suppliers to supply the exact number of units required to fulfill the order. The goods, however, are shipped to the etailer's site for packaging before they are finally delivered to the etailer's customer. This arrangement is currently being adopted by Amazon.com, to maintain quality control over packaging and shipping of products. This is especially so when multiple product types are sourced from different 3rd parties.

Delivery: There are a number of options: the delivery function can either be undertaken by the etailer her-

self, or she may choose to engage a 3rd-party logistics provider (3PL) to perform the function. The etailer, however, also has the option not to deliver the order to the customer's site. Instead, the order can be delivered to a particular outlet, so that the customer who places the order can pick it up from that site. All of these options appear to be adopted in the etailing industry.

This paper will evaluate, using a combination of mathematical modelling and numerical experiments, the first two basic inventory strategy options open to online retailers. The third option is more complex, and we leave it for future research. The first two options are further broken down into three cases as follows: the etailer

Case 0: maintains her own inventory; or

Case 1: maintains zero inventory but co-operates with one retailer or

Case 2: maintains zero inventory but co-operates with two or more retailers who normally have their own customers in addition to the etailer's. For each customer order the etailer receives, she turns to a retailer to fulfill the order.

Examples of etailers practicing models of Cases 1 and 2 are Zappos.com and Spun.com. Zappos.com sells shoes online and orders are delivered directly from manufacturers' or distributors' warehouses. In some seasons, Zappos.com ships customer orders (via UPS) free of charge. Spun.com sells CD/DVD but holds no inventory. It allies with the wholesaler Alliance Entertainment Corp. which stocks the CDs/DVDs and ships them directly to Spun.com's customers with Spun.com's labels on the packages.

In this paper, we will focus our analysis on the above three cases. Under a given set of conditions, a mathematical model based on dynamic programming will be used to derive the expected profit for the etailer as well as all 3rd-party suppliers. We will compare the above three strategies from the etailer's viewpoint, using profitability as the basis for comparison. Numerical experiments will be conducted to determine conditions under which the etailer will be better off holding her own inventory, as compared to relying on the inventories of 3rd parties to fulfill her customer orders.

Based on the numerical experiments, we will also provide insights on key factors affecting the etailers' decision on inventory strategy. Such insights allow us to explain why major online retailers such as Amazon.com are deviating from their original strategy of zero inventory to the present decision to invest extensively in warehousing, from a purely inventory management point of view. The model also provides a framework for etailer startups or existing etailers to evaluate their inventory strategies.

The close global attention directed towards the development of electronic commerce has resulted in the proliferation of published information on this area. Numerous publications on all aspects of e-commerce appear every-

day, and surveys are actively conducted by numerous agencies. Academic-oriented journals dedicated to the subject of e-commerce include the *International Journal of Electronic Commerce* and *Journal of Organizational Computing and Electronic Commerce*. The academic-oriented research on electronic commerce to-date, however, is qualitative in nature and none of them has directly addressed issues related to the order fulfillment process of etailing. The literature to-date focuses largely on the marketing, information technology, and overall growth aspects of Internet retailing (e.g., Lei and Robey, 1999). Here we only briefly review those that are mostly related studies, and the reader can find a more comprehensive survey in Chen, Hum and Sim (2002).

In essence, our business model of outsourcing inventory by etailers is similar to the term of drop-shipping used in the marketing literature. However, the potential for drop-shipping has been deemed limited in many earlier marketing books (e.g., Scheel, 1990) until the recent Internet era, since the latter has made possible real-time data integration with the physical concept of drop-shipping the goods by third parties directly to the customers. In fact, as pointed out in a parallel work by Netessine and Rudy (2000) on supply chain structures incorporating drop-shipping, an Eetailing World survey of internet retailers show that 30.6% of Internet-only retailers use drop-shipping as a primary mode of order fulfillment. Netessine and Rudy focus, however, on Internet-based supply chain structures, comparing and evaluating coordination schemes for such chains rather than on the operational aspect of inventory management of etailers.

Despite the explosive growth of information on e-commerce in general and Internet retailing in particular, it can be seen from the above literature review that currently there is not much literature which addresses the order fulfillment issues of online retailers through mathematical modelling. This paper therefore seeks to make a contribution by sketching out the operational issues involved in the order fulfillment process, and then focusing on one particular issue for a start, namely that of inventory decisions for the online retailer. The inventory literature related to the mathematical model(s) to be used in this study will be detailed in Section ??.

The rest of the paper is organized as follows. The mathematical model to assess the various inventory strategies of an etailer will be sketched in Section 2, and numerical experiments and analysis will be reported in Section 3. The summary conclusions of this study will be presented in Section 4.

2 Models

In this study, we will consider a single product, newsvendor type situation. For any retailer (etailer) who holds inventory, orders placed by him (her) will arrive at the beginning of the finite sales season, and there will be no additional replenishment during the season. Such a sce-

nario is representative of many real-life situations such as in the cases of news vendors, booksellers, and other seasonal products, which are all equally applicable to eetailing.

Besides physical retailers, independent 3rd-party suppliers for Case 1 may also be manufacturers, wholesalers, or distributors. In general, there are more physical retailers than wholesalers for any given product in any part of the world. As such, co-operating with bricks-and-mortar retailers gives the etailer the opportunity to gain an extensive network of geographically-dispersed stocking points near her customers, without having to set up her own warehousing infrastructure. The potential benefits of such an arrangement to the etailer are obvious: lower delivery costs and shorter delivery lead-time with minimum investments in warehousing and distribution infrastructure.

One of the major concerns of such an arrangement, however, is that retailers tend to give priority to their own customers (including their own online customers). In general, retailers' own customer orders are easier to handle, while those from the etailer are normally small and arrive more frequently and hence more troublesome to fulfill. Because of this, even manufacturers, wholesalers, and distributors may not be so responsive to etailers' orders as to physical retailers'.

Clearly, when a retailer agrees to co-operate with an etailer as in Case 1, both parties would have agreed on some form of "profit sharing" arrangement. For instance, the retailer could give the etailer a discount off his normal retail price for every order he receives from his online counterpart. Hence, by agreeing to co-operate with the etailer, the physical retailer effectively enters into a situation whereby he faces two classes of demand: those from his own customers (the high-profit customers) and those of the etailer (the low-profit ones). If the physical retailer's shortage cost for failing to fulfill his electronic counterpart's orders is less than the shortage cost to his own direct customers, any retailer who seeks to maximize his own profit will naturally find it more attractive to sell to his own customers rather than those of the etailer.

When replenishment is not possible during the sales season, whenever the physical retailer receives an order from his online partner, he faces the decision of whether to satisfy the demand or reject it so as to reserve his limited pool of inventory in the hope that there will be demand from his own customers later on. Given such a situation, the online retailer - who does not have control over the physical retailers' stocking policies - becomes vulnerable and her reliability to her own customers may very likely suffer.

For Case 2, we first consider the option of outsourcing to only two retailers, and then extend the analysis to the case of "infinitely" many retailers. This last special model will provide an upper bound on all outsourcing strategies in terms of profitability. Such retailers can be upstream suppliers as well, such as manufacturers, distributors, wholesalers, etc, who hold inventory. Sim-

ilarly, these upstream suppliers also tend to give higher priority to demands from other sources, such as those from their own on-line operations and bulk orders from other retailers. Nevertheless, our modelling approach and numerical results are also applicable even when 3rd-party suppliers give etailers' orders an equal priority, in terms of order fulfillment from their inventory.

We describe the above structural scenarios because in the lesser developed economies, many retailers are not ready to sell through the internet. In such a context, an etailer comes onto the scene to capitalize on the existing network of physical retailing outlets and therefore enter into alliances with these retailers in the manner highlighted above. Also, because of the lack of ready transportation for the individual consumer, the etailer is able to reach a market not often available to a given retailer. Because of this, retailers are willing to link up with the etailer.

As detailed in Section ??, besides inventory strategy, the etailer order fulfillment process comprises two other key elements: packaging and delivery. In this study, we assume that the etailer will let a 3rd party perform packaging and delivery functions in all three cases. The cost of packaging and delivery will be borne by buyers (as in Amazon.com which charges customers for these costs). To make the analysis as simple as possible, we ignore fixed-investment related costs such as rent/warehousing, IT, etc. Nonetheless, the assessment method to be developed here gives the etailer a guide as to the amount of fixed investment (as well as how much packaging/shipping costs) she can afford to absorb.

The further general assumptions made for all three cases are as follows:

- (i) For conciseness, we assume the sales season is divided into time intervals so short that the probability of more than one customer arriving in any interval is negligible.
- (ii) The probabilistic rule governing the arrival patterns of all customer types is assumed to be known, and every arrival will result in an order of one unit only.
- (iii) Initial inventory is zero and backlogging is not permitted - unmet demand will be lost. (This is the scenario we chose to model initially for this study).
- (iv) Demands for retailer(s) and the etailer are independent of each other. (In some context, as indicated earlier where accessibility to a given retailer is limited, this assumption may be reasonable).
- (v) Holding cost is incurred for all units unsold at the end of the sales season; if it is negative, then it represents the net salvage value.
- (vi) The prices charged to consumers by the etailer and each physical retailer are constant and known.
- (vii) The cost of goods, regardless of whether it is incurred by the etailer or any of the retailers, is constant and known for any order quantity.

- (viii) The sales season is short, such that the time value of money is negligible.

In Chen et al. (2002), we have formulated all mathematical models for all the cases with the above assumptions. These models turn out to be close the model that deals with optimal rationing policies by Gerchak, Parlar and Yee (1985, GPY hereafter). GPY consider a situation in which two distinct classes of customers with different per unit profit contributions exist for a single product or service in a fixed, finite sales season. The main objective of their study is to derive optimal rationing policies at any time (i.e., whether to reject those “low-revenue customers” upon arrival) depending on the number of units on hand and time left until the end of the sales season.

The model by GPY assumes that every arrival will result in an order of one unit only (as we do here). Banerjee and Viswanathan (1989) extend the study of GPY by considering Poisson batch-sized demand. They demonstrate using numerical examples that there does not exist a simple optimal rationing policy. According to McGill and Van Ryzin (1999), the model with batch demand under the yield/revenue management context remains an open problem. The models addressed in the revenue management literature are generally more complicated as they deal with both optimal pricing and inventory allocation among multiple classes of demands. However, that literature typically assumes the total number of inventory units is fixed.

3 Numerical Experiments and Discussions

The objective of the numerical experiments is to assess and compare the profitability of the etailer under each of the three cases detailed in Section 2.

Steps Taken in the Numerical Experiments

Adopting all the assumptions spelt out in this paper, we assess the three cases at different levels of all key parameters in the model. They include:

- Demand rate: we will assess the impact on all parties’ profitability when the etailer’s demand is significantly higher/lower than that of the retailers.
- Profit margin: we will change per unit product cost to evaluate the impact of profit margin and vary the “commission” rate that the physical retailer(s) gives to the etailer to see the relative impact of profit margins.
- Holding cost: the optimal ordering quantity is influenced by the holding cost.

- Shortage cost incurred by retailer(s) in failing to fulfill the etailer’s order: this affects the retailer’s rejection decision and subsequently the profitability of both parties.

We compare how each of the cases performs under different scenarios.

Brief Summary: Three key insights have been drawn from the modelling work and experiments. First, the numerical results suggest that the online retailer is better off relying on others to fulfill her orders when she faces small demand in the market (when her demand is relatively low compared to the physical retailers). When she forms alliances with 3rd party(ies), she should go with more than one supplier. As the online retailer’s sales volume becomes sufficiently large, she should revert to the strategy of maintaining her own inventory. With a high demand, the risk of having unsold units at the end of the selling season decreases. Also, by carrying her own inventory, the online retailer is able to enjoy a higher profit margin for every unit she sells, since she no longer needs to share the profit with 3rd parties as in the case of relying on others to fulfill her customer demand.

Second, when the profit margin is relatively large and retailers offer deep discounts to the etailer, the latter would be better off by outsourcing order fulfillment rather than holding her own inventory to meet orders. Third, with regard to holding cost, the outsourcing strategy will be attractive to the etailer if the holding cost is relatively high. Several outsourcing strategies were evaluated numerically based on heuristic solutions.

These findings confirm the general trends we observe in the real world of online retailing. For instance, major etailer Amazon.com started off with the strategy of zero inventory and heavy reliance on 3rd parties for order fulfillment. But as the online bookseller’s business grew, it deviated from this original strategy to one that involves extensive investments on warehousing and distribution infrastructures. Though many other factors may account for their strategic direction and practices, one cannot underestimate the underlying driving forces highlighted in this study. In another example, a MOR (maintenance, operations and repair) B2B etailer AsiaEC.com holds stocks of office furniture and PC related product categories, while it outsources other products to wholesalers and other MOR stores.

Further Managerial Implications: As an etailer’s business grows, the option of holding her own inventory becomes increasingly attractive. But instead of investing in his own inventory, the online retailer can consider the option of adopting Vendor Managed Inventory (VMI) program, which has been receiving close attention in the physical retailing scene during recent years. Under VMI, the etailer does not take ownership of the inventory until it is sold.

But regardless of whether the online retailer adopts VMI or decides to invest in his own inventory, warehousing

and distribution are still essential components of the etailer's fulfillment process. The etailer, however, may subcontract this function to a 3PL. Some of the US trucking and logistics firms, such as Schneider National, are building dedicated business-to-business e-commerce fulfillment operations. When a buyer clicks on the "ship now" button on paperloop.com, a bulk paper distributor, the order goes directly to Schneider, which fills it from its own warehouses onto its own trucks.

Clearly, the etailer can undertake a combined strategy of adopting VMI and engaging a 3PL for its warehousing management and distribution functions. By undertaking this combined strategy, the online retailer thereby frees himself once again to concentrate on the core activity of developing the online business.

4 Concluding Remarks

We have confined the analysis of inventory strategy within the context of standard inventory costs, and other factors such as packaging, competition, multiple items and order consolidation, gaming, and fixed investments have not been touched in this study. Clearly, these suggest further research possibilities.

In our model, the hybrid inventory strategy of both holding and outsourcing inventory is not addressed. This is alike the inventory transshipment model, and therefore, the literature on transshipment models may serve as an excellent starting point (see Rudi, Kapur and Pyke, 1998, for this). This paper considers inventory decisions, and the models developed in this study permit systematic evaluation of various options.

The demands among retailers (including the etailer) have been assumed independent of each other. While we have indicated possible scenarios where this assumption may be reasonable, it is clearly appealing to extend the analysis to the case where demands among retailer(s) and the etailer are interdependent or that demand cannibalization exists. However, this calls for a substantially different modelling approach as the current model is incapable of handling the demand correlation.

As our models are based on the assumption that no more than one unit of demand arrives at any time interval, an interesting area for future research is to consider batch arrivals. For instance, if a batch order is placed by the online retailer, should the physical retailer reject the entire batch order or fulfill it partially? Also, does it make sense for the online retailer to allow the physical retailer to fulfill the order partially if the batch order originates from a single customer? Finally, if the online retailer does not allow partial fulfillment of orders, will the physical retailer be better off accepting the batch order at the expense of rejecting his own direct customers? Though some of these issues are not unique to the Internet context (e.g., also raised in GPY), they are interesting issues which future research may wish to address.

To ensure a reliable supply, the etailer may resort to other leverages such as commitment contracts with 3rd parties. One scenario would be that the etailer commits to a minimum quantity and the 3rd party guarantees a minimum service level. The research along this line is under way.

Quite clearly, this study is aimed at delineating the issues involved in the entire online order fulfillment process, and our focus on inventory decisions is just a first modelling-based attempt at contributing to this new and rich area of research.

5 Acknowledgement

This research is partially supported by Hong Kong RGC/CUHK Grant 2050271, Grants R-314-000-025-112 and R-314-000-002-112 from the National University of Singapore.

References

- [1] Banerjee, P. K. and B. Viswanathan. 1989. On Optimal Rationing Policies. *Canadian Journal of Administrative Science*, 6 (4), 1-6.
- [2] Chau P.Y.K., G. Au and Y.T. Kar. 2000. Impact of Information Presentation Modes on Online Shopping: An Empirical Evaluation of a Broadband Interactive Shopping Service. *Journal of Organizational Computing and Electronic Commerce*, 10 (1), 1-22.
- [3] Chen, Y., S. H. Hum, and H. Sim. 2002. On Inventory Strategies of Online Retailers. Unabridged version of this paper, available at www.se.cuhk.edu.hk/yhchen.
- [4] Gerchak Y., M. Parlar and T.K.M. Yee. 1985. Optimal Rationing Policies and Production Quantities for Products with Several Demand Classes. *Canadian Journal of Administrative Science*, 2 (1), 161-176.
- [5] Lei J. and D. Robey. 1999. Explaining Cybermediation: An Organizational Analysis of Electronic Retailing. *International Journal of Electronic Commerce*, Summer, 3 (4), 47-65.
- [6] Lippman, S. A. and K. F. McCardle, 1997. The Competitive Newsboy, *Operations Research*, 45 (1), 54-65.
- [7] McGill, J. I. and G. J. Van Ryzin. 1999. Revenue Management: Research Overview and Prospects. *Transportation Science*, 33(2), 233-256.
- [8] Netssine, S. and Nils Rudi. Oct., 2000. Supply Chain Structures on the Internet: Marketing-Operations Coordination. Working paper, W.E. Simon Graduate School of Business Administration, Univ. of Rochester.
- [9] Ricker F.R. and R. Kalakota. 1999. Order Fulfillment: The Hidden Key to E-Commerce Success. *Supply Chain Management Review*, Fall 1999, 60-70.
- [10] Rudi N., S. Kapur and D. F. Pyke. 1998. A Two-location Inventory Model with Transshipment and Local Decision Making. *Working Paper*, The Amos Tuck School, Dartmouth College, Hanover, NH 03755.
- [11] Scheel, N. T. 1990. Drop Shipping as a Marketing Function. Greenwood Publishing Group.

On-Demand Business Process Integration Based on Intelligent Web Services

Xiaohua Lu¹, Yinsheng Li¹, Ying Huang²

¹Software School, Fudan University, Shanghai, China

Phone: +86-21-55664096-808, {0014010, liys}@fudan.edu.cn

²IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, USA

yxh@us.ibm.com

ABSTRACT

This paper proposes to make On-Demand business process integration based on intelligent Web Services. Business processes are modeled with BPEL4WS. They are incorporated into intelligent Web Services as their knowledge. Implementation framework and techniques are presented. A case of online auction is developed to demonstrate these concepts.

Keywords: on-demand business, process template, intelligent Web Service, BPEL4WS, online auction

1. INTRODUCTION

Businesses are becoming customer-centric from product-oriented. Enterprises have to respond with speed to customer demands, market opportunities or external threats. They therefore develop their e-Business systems to be more responsive, variable and resilient, i.e., on demand. To this end, their electronic business processes are required to integrate end-to-end across the company and with key partners, suppliers and customers. Service-oriented architecture (SOA) provides a competent computation environment for enterprises to integrate their applications and become on demand. Businesses can become flexible through integration of people, processes and information within and beyond enterprises. SOA and standardized Web services technologies make it possible to simplify information systems through automation and virtualization, enable access to and create a consolidated, logical view of resources across a network. Every application and resource is represented as a service with a standard interface which simplifies its activation and the exchange of information. Enterprises can treat business processes and the underlying infrastructure as defined components that can be mixed and matched at will.^[4, 5]

To this end, numerous research efforts have been put on SOA based business process integration. This paper describes process template-based business integration, which is an ongoing research at IBM Watson Research Centre. We locate characteristics and advantages of template-based on-demand business, and try to improve it with intelligent Web services, which was proposed and studied in our previous publications.^[1] The proposed intelligent Web services incorporate possible business processes, which are modeled with BPEL4WS, into its knowledge. They run as composition services and flexible business adapters in SOA-based application integration environments. The knowledge makes them rational to response to on-demand business or transaction requests with appropriate processes and call appropriate services. This will be demonstrated through a case study

of online auction for e-Marketplaces, in which an auction service is implemented with intelligence and can instantiate an appropriate process based on the requester's inputs, to accomplish an auction session.

2. TEMPLATE-BASED ON-DEMAND BUSINESS INTEGRATION

Today's e-business systems have been found with several drawbacks. An e-Business solution is usually created for a single enterprise, not flexible for an extended business. Systems of suppliers, manufacturers, sellers and customers are all developed with different technologies, running on different platforms. It is very difficult and costly when integrating systems of the whole supply chain. Also, it is not able to handle changes quickly and effectively when being required to meet requirements by another enterprise because business processes are buried in program codes. When processes changes, all these codes have to change. This is time-consuming and may lead to loss of potential business opportunities. Under on-demand business environment, all system components are services, and the whole system is constructed with service-oriented architecture. Some basic services are provided by platforms and service providers, while other business specific ones are created by enterprises own. That is, enterprises only need to focus on their business issues. We can create e-business solution by developing service components and defining templates to integrate these services. With this method, we can integrate e-business systems of different partners and create business solutions easily. When defining templates, business partners across supply chain are all incorporated. As long as they provide corresponding services, business process can be easily integrated with fewer costs than before. All these services and templates are reusable components. Moreover, business processes are no longer buried in codes. We rewrite those templates rather than the entire application to meet process changes.

In template-based on-demand business integration, there

are two concepts with particular importance, which are services and templates. Services comprise of business services and IT services. Business services can be further categorized into four types: demand services, such as sales; supply services, such as logistics; administrative services, such as human resource; control services, such as finance. IT services also have four categories: adapter services, such as data adapter; composition services, such as process flow; platform services, such as messaging; common services, such as authentication and authorization. Among these services, some can be found at service providers, while others need to be developed by enterprises. [2, 3]

When enterprises have all service components for e-business systems, they define business templates to compose those services into business processes and applications. Templates describe how services work with each other and how business processes execute. Composition services are used to integrate other services according to templates. Figure 1 depicts the framework of e-business solution using services and templates.

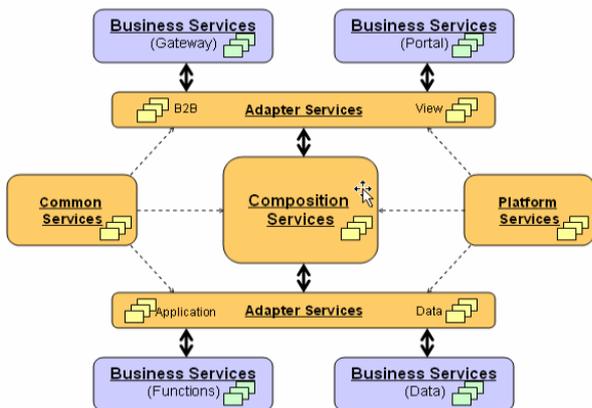


Figure 1. Framework of e-business solution using services and templates [2, 3]

In this framework, business services provide business functions. Common services and platform services provide facilities required by every application. Adapter services facilitate interaction between business services by providing agreed interfaces for them. And composition services integrate all the services above to form process templates and e-business system.

Template-based e-business solution has following advantages:

1. Process templates are independent of technologies and platforms. Changes in them will not impact existing system.
2. Templates and services are all individual artifacts which can be stored and reused later;
3. Templates and services are conformed to standard interfaces. They can be reassembled and reconfigured on demand.

While creating e-business solutions using process templates, we first identify required business services

and IT services. Secondly we search service directory for desired services. We hire them if existed. Otherwise we have to create by our own or request them from other sources. The following step is to define business templates according to business processes, which integrate the services above. Finally, we use these templates to accomplish service composition and create complete e-business solution. The following figure delineates the entire process.

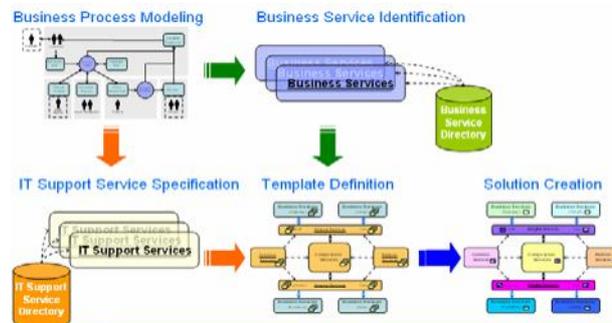


Figure 2. Building e-business solution through template and composition [2, 3]

When business process changes, enterprise only need to develop new service components and define new templates. New application can be easily integrated. Compared to rewriting the whole application, this template-based method reduces cost and makes enterprises more responsive.

3. PROCESS MODELING BASED ON INTELLIGENT WEB SERVICE AND BPEL4WS

More and more enterprises are accepting web-based e-business systems. Web services have been envisioned to be one of the key integration technologies. We have enhanced Web services with intelligence-enabled internal model. Semantic Web Service technologies have not been mature enough to represent knowledge. [8] We therefore incorporate well-developed agent knowledge representation model with Web services internal entities. Such intelligent Web services can contribute to the process composition automation, as well as interaction with partners' systems within on demand business solutions.

BPEL4WS (Business Process Execution Language for Web Service) is an XML-based language used to integrate Web Services to model process. [10] That is, it creates templates to compose Web Services. What created by BPEL4WS could also be a Web Service and described by WSDL. BPEL4WS defined how Web Services involved in the process cooperate and interact with each other to achieve business Goals. To implement on-demand business, we can define several business processes using BPEL4WS, and a composition Web Service will use these templates to invoke other Web Services to accomplish the whole business. Such composition Web Service is implemented using

intelligence and has several business processes modeled with BPEL4WS as its knowledge. It chooses the most appropriate process according to external events (e.g. user request) and internal rules, and executes the selected processes to realize on-demand business. Figure 3 illustrates the architecture of process modeling based on intelligent Web Service and BPEL4WS.

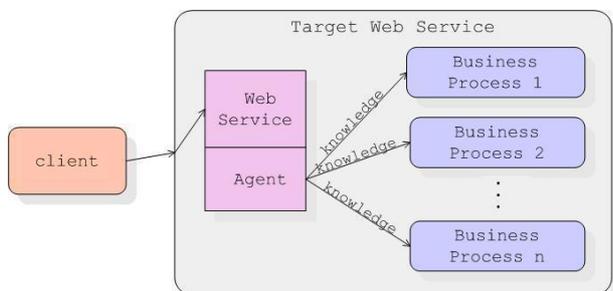


Figure 3. Architecture of process modeling based on intelligent Web Service and BPEL4WS

An end user issues business request through its client software (“client” in above figure), which sends this request to “Target Web Service”. “Target Web Service” is implemented with intelligence. It is developed as a software agent with standard Web Service interfaces. With this paradigm, the intelligent Web service can use its own knowledge representation model to incorporate business processes described by BPEL4WS as its knowledge, and then choose the most appropriate one to execute on environment changes and its intelligence. In above figure, from “Business 1” to “Business n” are all represented as agent-applicable knowledge.

4. IMPLEMENTATION ISSUES

For the mentioned case, we use software agent which conforms to FIPA (Foundation for Intelligent Physical Agents) standards to develop intelligent Web service.^[11] The services are implemented as agents and published as Web Services. FIPA agent uses ACL (Agent Communication Language) for communication, which is not compatible with Web Service’s communication protocol, SOAP. When using agent to implement Web Service, we must address this discrimination in communication, that is, to make agent understand SOAP. There are two solutions to this problem:

1. Enhance FIPA agent and make it understand SOAP. With this method, agent can communicate with each other using both ACL and SOAP. But it is too difficult since it is almost the same as developing one specific FIPA agent implementation. In addition to difference in communication protocol, which means we have to add parsing and encapsulating of SOAP to agent functions, communication styles also differ. Web Services support synchronous and asynchronous invocation, while FIPA agent can only be called asynchronously. This adds a great deal to the complexity of this method and finally make we give it up.
2. Add a mediator module for FIPA agent, which can

understand both ACL and SOAP. The mediator receives SOAP message from client, extracts request parameters, creates new ACL message, then puts those parameters into new message and finally, sends ACL message to agent. When agent receives ACL message, it uses built-in parameters to perform requested tasks and returns result to the mediator. On receiving ACL-based result, mediator performs another message format translation and sends SOAP-based result back to client. Then one transaction completes. This method is much easier to implement than the previous one. So it is chosen in this paper.

With regard to how to incorporate BPEL4WS-based business process as agent’s knowledge, there are also two considerations:

1. Store BPEL4WS files, WSDL files and other assisting files somewhere FIPA agent can access (e.g. file system). When agent activates, it will go to the specified place, read those files and execute processes described in them. On doing so, it is required that agents must understand BPEL4WS and WSDL, as well as perform corresponding operations. Actually, it means adding BPEL4WS engine to agent, which is even more difficult than adding SOAP engine to agent. So it is discarded.
2. Make use of existing BPEL4WS engine and combine it with agent in some way. We choose to deploy BPEL4WS processes on BPEL4WS server and publish as Web Services. Then agent is able to invoke these Web Services with standard interface, as well as accomplish business tasks. When compared to previous one, this method is more convenient and flexible. Therefore, it is the better choice.

5. CASE STUDY: ONLINE AUCTION

For case study, we use the proposed architecture to implement an online auction business process. Customers can issue an auction request on client interface. Then the request is sent to the auction Web service. The service has a few BPEL4WS business templates as its knowledge and knows which one to choose. Then the service invokes the most appropriate BPEL4WS Web service. The selected process is instantiated as an auction session.

Following demonstrate those two business processes involved. One is called “OnlineAuction” and the other “OnlineAuction2”. These two processes are identical in the first four steps, described as follows:

1. Customer logins into auction web site, and then enters information about auctioned commodity and criteria used later to evaluate auction result.
2. Auction mediator receives such information and asks buyers to bid.
3. In this case there are two buyers. On receiving auction request, they will bid separately based on the

information customer entered.

4. Auction mediator collects prices from all buyers, calculates their scores, and determines a winner.

In “OnlineAuction”, the step 5 is:

5. Auction mediator returns price and score of winner to customer.

In “OnlineAuction2”, the step 5 is:

6. Auction mediator returns prices and scores of all buyers to customer, and then suggests a winner.

The differences between the two processes are that “OnlineAuction” only returns information about the winner; therefore auction decision is made by the mediator. While in “OnlineAuction2”, mediator returns information about all buyers and suggests a winner, but final decision is still up to the customer.

In this case, we choose a FIPA agent implementation, JADE (Java Agent Development Framework) to develop intelligent Web services. [9] In order to do so, we also use an open source project called “Wsag”. [6] It acts as a mediator between the Web Service and JADE agent and makes message format translation. “Wsag” is deployed on Apache Tomcat 4.1.24, while agent is running in JADE 3.1 environment. When publishing BEPL4WS processes as Web Services, we choose an existing BEPL4WS engine, BEPL4WS server 2.0 from collaxa. [7]

The following figure illustrates architecture of online auction case.

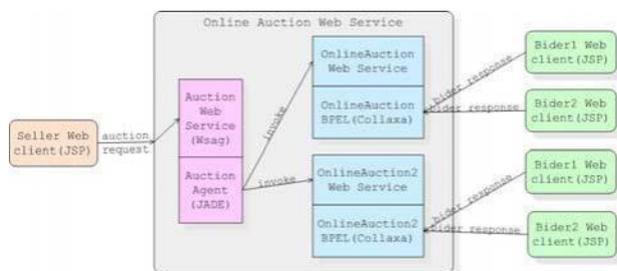


Figure 4. Architecture of online auction case

Following is a complete instance of online auction case. Customer visits auction web site (<http://10.89.103.1:8080/wsag/OnlineAuctionWebClient.jsp>), enters name and description about commodity to be auctioned and specifies weights for buyer’s price and reputation. The customer has to choose whether he wants to see bid information of all the buyers or only the winner. In this instance, “Name” and “Description” are set to “Astrology Mythology” and “A Very Interesting Astrology Book”, separately. The weight for “Price” and “Reputation” is 10:1. At last, “Only Return Final Result” is selected, i.e., the customer only wants to see bid information of the winner. This customer request is sent to Auction Web Service, which is implemented with agent and uses “Wsag” as mediator. Auction Agent which accomplishes actual business function is running

on another computer (auction0@SSC036:1099/JADE@<http://10.89.103.36:7778/acc>). When the agent receives request, it knows which process to invoke. If “Only Return Final Result” is selected in this instance, it will invoke OnlineAuction Web Service; otherwise OnlineAuction2 Web Service is called. These two Web Services are implemented with BEPL4WS and deployed on BEPL4WS server running on a third computer (<http://10.89.150.150:9700/collaxa/default/OnlineAuction> and <http://10.89.150.150:9700/collaxa/default/OnlineAuction2>). As aforementioned, these two processes involve human interactions. Buyers can login to web site to participate in these processes (<http://10.89.150.150:9700/OnlineAuctionUI/> and <http://10.89.150.150:9700/OnlineAuction2UI/>). A buyer named “bider1@software” has already login to OnlineAuction web site and sees all process instances waiting for his participation. Then he can examine information of auctioned commodity and criteria, as well as offer price and reputation. After collecting bids from all the buyers, process instance can go on. When calculating scores, process uses following formula:

$$Score = Price * PriceWeight + Reputation * ReputationWeight \quad (1)$$

The buyer with the highest score is the winner. Then process will return information of the winner to customer. If customer doesn’t choose “Only Return Final Result” at the beginning, process will return information of all the buyers and suggest a winner. Figure 5 shows snapshots for this auction instance.

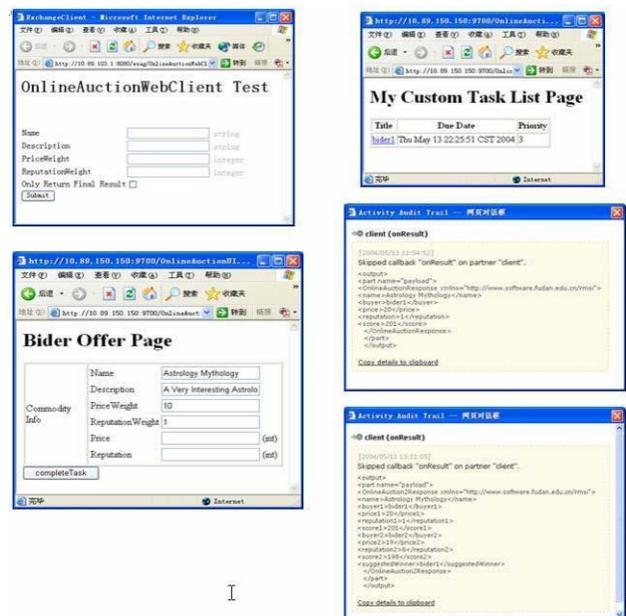


Figure 5. One instance of online auction

In this case, our proposed architecture has been proven feasible. However, Web Service’s intelligence is not exploited enough. This needs to be improved.

6. CONCLUSION

This paper proposes that business processes modeled

with BPEL4WS can be made part of intelligence Web service's knowledge to accomplish process modeling, as well as on-demand businesses. Several critical issues have been addressed, including system architecture, how to implement intelligent Web Service with agent-oriented knowledge, and how to represent BPEL4WS processes as knowledge. However, there are several issues that still need improvement, i.e., intelligent Web services themselves don't understand BPEL4WS without existing BPEL4WS engine, and we have not utilized intelligence well to make Web Service intelligent. These problems call for research efforts.

REFERENCES

- [1] Yinsheng Li, Hamada Ghenniwa, Weiming Shena, "Agent-based Web Services Framework and Development Environment"
- [2] Ying Huang, "e-business Solution Templates based on Web Services & Business Processes"
- [3] Ying Huang, "Model-driven integration for service-oriented on demand business solutions"
- [4] Irving Wladawsky-Berger, "Meet the Experts: Irving Wladawsky-Berger on e-business on demand", http://www-900.ibm.com/developerWorks/cn/dmdd/library/techarticles/0302iwb/0302iwb_eng.shtml
- [5] Alfredo Gutierrez, "e-business on demand: A developer's roadmap", http://www-900.ibm.com/developerWorks/cn/ebod/i-ebodov/index_eng.shtml
- [6] Whitestein Technologies, Wsag project, <http://wsai.sourceforge.net/gateway.html>
- [7] Collaxa Company, Collaxa home page, <http://www.collaxa.com/>
- [8] DAML organization, DAML Services home page, <http://www.daml.org/services/owl-s/>
- [9] Tilab, JADE home page, <http://jade.tilab.com/>
- [10] IBM, Microsoft, BEA, "Business Process Execution Language for Web Services Version 1.1 Specification", <ftp://www6.software.ibm.com/software/developer/library/ws-bpel.pdf>
- [11] FIPA organization, FIPA home page, <http://www.fipa.org/>

A Two Phase Verification Algorithm for Cyclic Workflow Graphs

Yongsun Choi

Dept. of Systems Management & Engineering, Inje University, Kimhae, 621-749, Korea
yschoi@inje.ac.kr

ABSTRACT

The widespread automation of e-business processes has made workflow analysis and design an integral part of information management. Graph-based workflow models enables depicting complex processes in a fairly compact form. This free form, on the other hand, can yield models that may fail depending on the judgment of the modeler and create modeling situations that cannot be executed or will behave in a manner not expected by the modeler. Further, cycles in workflow models needed for purposes of rework and information feedback increase the complexity of workflow analysis. This paper presents a novel method of partitioning a cyclic workflow process, represented in a directed graph, into a set of acyclic subgraphs. This allows a cyclic workflow model to be analyzed further with several smaller subflows, which are all acyclic. As a convincing example, we present two-phased verification of structural conflicts in workflow models for those incurred from the inappropriate composition of partitioned flows and the others within each acyclic subgraph, which is much easier to comprehend and verify individually than the whole workflow model, in general.

Keywords: workflow model, directed graph, feedback, partition, verification

1. INTRODUCTION

The recent surge of e-business research and development has resulted in the automation of thousands of business processes by means of workflow management systems (WfMS), both within and across corporate boundaries ([5], [7], [9], [16], [18], [30], [32]). WfMS enable the design, analysis, optimization, and execution of business processes. The basic WfMS functions include the separation between the business process logic and business applications, management of relationships among process participants, integration of internal and external process resources, and monitoring and control of process performances ([24], [34], [35]). There are typically two stages of workflow management, workflow specification and workflow execution. The former defines a workflow model, and the latter generates workflow instances guided by the workflow model.

A workflow model (also called workflow definition) consists of a set of activities and their execution sequences, the start and termination of the process, and the resources needed for executing individual activities, such as participants, associated IT applications, and data ([34], [35]). The objective of workflow modeling is to provide high-level specification of processes that are independent of the implementation intricacies of the target workflow management system ([8], [13], [18], [21], [22], [35]). Of all workflow perspectives, e.g., control-flow, data, organization, task, and operation, the control-flow perspective is the most prominent one because it defines the backbone of the workflow on which other perspectives can be specified [1].

Workflow models must be correctly defined before being deployed in a workflow management system to avoid any costly maintenance delays due to runtime errors in the process model [29]. Therefore, it is essential to verify

the workflow model before its deployment. Despite the importance of workflow verification, few commercial workflow systems provide formal verification tools. This lack of verification support can be attributed to the fact that most of the more than 250 commercially available WfMS use a vendor-specific ad-hoc modeling techniques [5] without theoretical framework for the representation, analysis, and manipulation of workflow systems [8]. Although these vendor-provided tools facilitate workflow specification and visualization, and some support simulation of processes under various conditions, they usually do not support formal workflow analyses. While process simulation can provide useful insight into process behavior, it does not address questions about the interrelationships among process components [8]. Consequently, few workflows are thoroughly checked before they are deployed in practice, often resulting in errors that need to be corrected in an ad hoc fashion at prohibitive costs.

There are two research approaches to ensure control flow correctness in workflow models – *build it correctly*, or *check it completely* [31]. The former ([10], [17], structured model in [21]), relying on strict rules in composing the model, may not model certain processes due to syntactical restrictions [21] and may not be very suitable for practical implementation in industry [31]. The latter, like Petri nets ([1], [4], [6], [26], [31], [33]) or the graph-theoretic techniques ([8], [14], [25], [29]), on the contrary, appeals more by allowing the user tremendous flexibility in expressing process requirements offering interesting analysis challenges to the modeler [31]. The Petri-nets-based workflow verification depends on the formalism of Petri nets, which is rarely adopted by existing WfMS, and makes it hard to be embedded into most of WfMS. Therefore, for non-Petri-nets workflow models, it is difficult to apply Petri-nets-based verification techniques as it requires the

transformation from each different modeling paradigm into the Petri nets formalism plus an effort of determining the corresponding structural conflicts in the original model. The majority of business processes have been analyzed using acyclic (i.e. loop-free) free-choice nets, a special class of Petri nets that enjoys the added advantage that *soundness* can be verified in polynomial time [1]. Additionally, establishing soundness in a free-choice net implies that the net is free from deadlock, and is also alive, i.e., no dead tasks [2]. However, the modeling of exceptions or precedence partially destroys the free-choiceness of the equivalent Petri net mapping. Moreover, modeling iteration necessitates the presence of loops in the control flow model, a problem that is yet to be satisfactorily addressed [31].

Graph-based workflow models have been widely used in industry to specify control flows of a business process. Although graphs provide a great flexibility for depicting complex processes in a fairly compact form, their free-form nature also needs to resolve models that may fail due to a lack of support for the modeler to avoid situations that cannot be executed or will behave in a manner not expected by the modeler [12]. Graph reduction ([25], [29]) and block-wise abstraction [14] have been proposed to identify structural conflicts in workflow graphs, but both approaches are limited to acyclic models [3].

This paper presents a novel method of partitioning into a set of acyclic subgraphs and decomposed verification of structural conflicts for a cyclic workflow process, even for nested structures, represented in a directed graph. Our approach of feedback partitioning and decomposed verification is unique and has several distinctive features useful for workflow analysis and design. First, we present two-phased verification of structural conflicts in workflow models. Our method detects certain types of structural conflicts, in the early stage, incurred from the inappropriate composition of feedback flows. And then, our method applies verification to each acyclic subgraph, which is much easier to comprehend and verify individually than the whole workflow model, in general. This decomposition feature has not been seen in any existing workflow verification method. Second, each acyclic flow can be utilized in hierarchically decomposing the given model and is a candidate for a sub-workflow, the knowledge of which is helpful for managing large-scale workflows. Finally, our method is based on a directed graph representation, more generic and adopted widely by the WfMS vendors, and utilizes well-known algorithms of graph-theoretic techniques, which makes it easy to embed our approach into most WfMS.

The rest of the paper is organized as follows. In Section 2, we present the preliminary concepts such as directed graph representation and types of structural conflicts of a workflow model. In Section 3, we present the way of partitioning a cyclic workflow graph into the

subgraphs of acyclic flows with illustrative examples. Section 4 delineates the two-phased verification of structural conflicts and presents the associated theorems. Section 5 discusses related work, summarizes our contributions, and outlines future research directions.

2. WORKFLOW MODELS IN DIRECTED GRAPHS

A workflow graph is a directed graph $G = (N, T)$ with a set of nodes N and a set of edges $(i, j) \in T$, where $i, j \in N$. Each edge, called a *transition*, links two nodes and represents the execution order of nodes. A *node* is classified into two types, task and coordinator. A *task*, represented by a rectangle, stands for a unit of work to be done. A *coordinator*, represented by a circle, stands for a point where only one of succeeding paths is selected or several paths are merged. Depending on the types of nodes and the number of incoming and outgoing transitions, nodes can be classified into five categories, i.e., *sequence*, *AND-split*, *AND-join*, *XOR-split*, and *OR-join*, as shown in Fig. 1. *Start* and *End* nodes are used to indicate the beginning and the end of the given workflow process, respectively ([25], [29]).

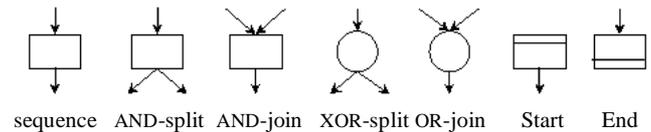


Fig. 1. Classification of nodes in workflow graphs
Fig. 2(a) illustrates an example cyclic workflow, extending an acyclic example in [29]. We make a simplifying assumption on the workflow graph that a node cannot be a join and split at the same time. Any node that joins and concurrently splits can be converted into a set of a join and a split, separate and connected with a transition.

Structural conflicts in workflow models

In this paper, we focus on the verification of three types of structural conflicts, deadlock, lack of synchronization, and livelock as illustrated in Fig. 3 ([1], [20], [29]). *Deadlock* refers to a situation in which a workflow instance gets into a stalemate such that no further activity can be executed. This happens when only some partial subset of the join paths to an AND-Join is executed, the AND-Join node k will wait forever and block the continuation of the process (Fig. 3(a)). *Lack of synchronization* refers to a situation in which the concurrent flows are joined by an OR-Join, resulting in unintentional multiple executions of activities following the node k (Fig. 3(b)). *Livelock* refers to a situation in which certain loop(s) of tasks are continuously performed, and there is no execution path leading to termination or cannot terminate properly (Fig. 3(c)). Note that livelock occurs only where a cyclic structure exists.

Other types of structural conflicts such as *dangled* nodes are relatively easier to detect than those described above,

from the graph-theoretic viewpoint. A dangled

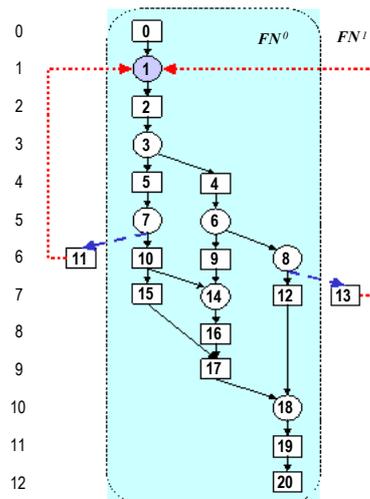
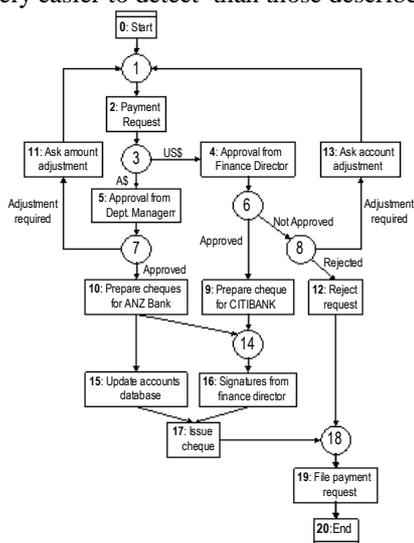


Fig. 2. (a) An example cyclic workflow graph; and (b) the normalized graph

node can be found simply by checking the reachability of each node from the Start node and to the End node.

discovered yet, when to explore edge (u, v) in the course of *Depth-First Search*. With all *back edges* removed, the rank of each node can be computed from the resulting acyclic workflow graph with the complexity of $O(|T|)$ [19]. Fig. 2(b) shows the normalized graph of the workflow process in Fig. 2(a), with nodes rearranged by its rank $r(i)$, $i \in N$, computed from the resulting acyclic workflow graph and shown to the left in Fig. 2(b). Fig. 4(b) shows another example of normalized graph for a workflow of Fig. 4(a) with nested cycles.

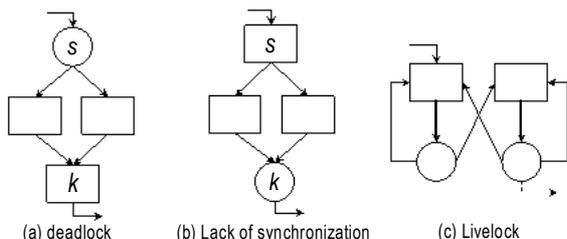


Fig. 3. Types of structural conflicts

3. PARTITIONING A CYCLIC WORKFLOW INTO ACYCLIC FLOWS

Cycles in workflow models are needed for purposes of rework and information feedback. For instance, the workflow in Fig. 2 contains two simple cycles, i.e., a cycle that does not contain any other cycles [19]. Our method identifies and partitions off each acyclic flow from the given cyclic workflow model. At the first iteration of partition, our method partitions off a maximally connected subgraph from the *Start* to the *End*, spanned by all nodes reachable to the *End* without multiple visits to any nodes contained, called *main flow*. If the original workflow model has nested feedback structures, our method proceeds on partitioning further to the remaining subgraph(s) that contain(s) cyclic structures. Otherwise, our method identifies and partition off each additional acyclic *feedback flow* from the remaining subgraph(s). More formal definitions and detailed descriptions of our method are given next.

The algorithm of *Depth-First Search* can be applied to identify the “back edges” where each of those completes a cycle, with the complexity of $O(|N|+|T|)$ [15]. An edge (u, v) is classified as a back edge if v is already discovered but any of its adjacent nodes are not

A normalized graph is helpful for identifying feedback structures of a cyclic workflow process. By examining the normalized graphs in Fig. 2(b) and Fig. 4(b), we find some interesting phenomena. First, every dotted upstream transition (i, j) , with $r(i) > r(j)$, merges a feedback flow, which is a back edge as defined above. Second, each dashed transition initiates a new feedback flow. Transition $(13, 7)$ in Fig. 4(b) has both characteristics. These transitions can be used as the cut sets to group the nodes in N . Next, we describe an iterative way of identifying those transitions initiating or merging each feedback flow and partitioning the graph of a cyclic workflow into subgraphs of acyclic flows.

Steps of a partitioning

The back edges of a cyclic workflow graph found during normalization are called as *Feedback Join Transitions (FJT)*. Each transition (i, j) of *FJT* has the property of $r(i) > r(j)$ and corresponds to a distinct simple cycle of a cyclic workflow graph. $FJT = \{(11, 1), (13, 1)\}$ and $FJT = \{(11, 7), (13, 7), (23, 19), (24, 1)\}$ for the examples shown in Fig. 2 and 4, respectively. Nodes that merge feedback flows are called as *Feedback Joins*, denoted by *FJ*, where

$$FJ = \{j \mid (i, j) \in FJT\}, \tag{1}$$

$FJ = \{1\}$ and $FJ = \{1, 7, 19\}$ for the examples shown in Fig. 2 and 4, respectively.

Definition 1. For any $i \in N$, let the *Order of Feedback* of i , called by $OF(i)$, denote the minimum number of transitions of FJT to pass to the *End* node. The subset of nodes with $OF(i) = n$ will be called as n^{th} -order *Feedback Nodes* and denoted by FN^n .

Note that $N = \cup_n FN^n$. Nodes of FN^0 do not need any transitions in FJT to reach the *End* node, and FN^0 can be identified as follows,

$$FN^0 = \{ i \mid R_{FJT}(i, End) = 1 \}, \quad (2)$$

where $R_{FJT}(i, End)$, which can be computed with complexity of $O(|T|)$, denote the *reachability* of node i to the *End* node without the transitions of FJT . That is, $R_{FJT}(i, End) = 1$ when node i can reach the *End* node without passing through any transitions in FJT , or $R_{FJT}(i, End) = 0$ otherwise. $R_{FJT}(End, End)$ is defined to be 1. Table 1 shows $R_{FJT}(i, End)$ for each node i , resulting $N - FN^0 = \{ 11, 13 \}$ for the workflow in Fig. 2. It can be shown that $N - FN^0 = \{ 11, 16, 17, 18, 19, 20, 21, 22, 23, 24 \}$ for the workflow in Fig. 4. The subgraph spanned by FN^0 will be called as the *main flow* and will be denoted as $MF(Start, End)$.

Table 1. $R_{FJT}(i, End)$ of each node i in Figure 2

i	0	1	2	3	4	5	6	7	8	9	16	11	12	13	14	15	16	17	18	19	20	
$R_{FJT}(i)$	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1

Definition 2. The n^{th} -order *Feedback Joins*, denoted by FJ^n , is the subset of feedback joins in FN^n . That is $FJ^n = FJ \cap FN^n$. The n^{th} -order *Feedback Splits*, denoted by FS^n , is the subset of nodes that split the subgraph spanned by FN^{n+1} from the subgraph spanned by FN^n , identified as $FS^n = \{ i \mid (i, j) \in T, i \in FN^n, j \in FN^{n+1} \cup FJ^n \}$.

Note that $FJ = \cup_n FJ^n$. By definition 2, we get FJ^0 as follows,

$$FJ^0 = FJ \cap FN^0. \quad (3)$$

$FJ^0 = \{ 1 \}$ and $FJ^1 = \{ 1, 7 \}$ for the examples of Fig. 2 and 4, respectively.

The set of nodes in any of FS^n will be called as *Feedback Splits*, denoted by FS , where $FS = \cup_n FS^n$. Finally, FS^0 will be identified. Since FN^1 is not available yet and there will be no transition from a node in FN^0 to any node in $\cup_{n \geq 2} FN^n$, we identify FS^0 as follows,

$$FS^0 = \{ i \mid (i, j) \in T, i \in FN^0, j \in (N - FN^0) \cup FJ^0 \} \quad (4)$$

$FS^0 = \{ 7, 8 \}$ and $FS^1 = \{ 10, 13, 15 \}$ for the workflow in Fig. 2 and 4, respectively.

At each iteration, our method identifies FN^n , FJ^n and FS^n . FN^n and FJ^n is identified by classifying N and FJ , respectively. On the other hand, FS^n is newly discovered from FN^n , at each iteration. When all the nodes in FJ are classified into one of FJ^n , we can conclude that $FN^{n+1} = N - \cup_{m \leq n} FN^m$ and no further classification of nodes is required. For instance, $FJ^1 = FJ$ for the workflow in Fig. 2, and we can conclude with $FN^1 = N - FN^0 = \{ 11, 13 \}$. Otherwise, there exist more feedback structures in

the subgraph(s) spanned by $N - \cup_{m \leq n} FN^m$ and further classification of nodes is required.

Recursive partition for nested feedback structures

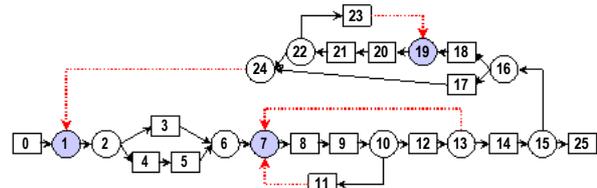
Definition 3. For any $fs \in FS^{n-1}$, let $Desc^0(fs)$ denote the set of nodes in FN^n and $Desc^{n+1}(fs)$ denote the set of nodes in $\cup_{m \geq n} FN^m$, respectively, that can be reached from fs by the transitions in T . For any $fj \in FJ^{n-1}$, let $FAncs^n(fj)$ denote the set of nodes in FN^n and $FAncs^{n+1}(fj)$ denote the set of nodes in $\cup_{m \geq n} FN^m$, respectively, that can reach fj by the transitions in T .

Definition 4. Let $fs \in FS^{n-1}$ and $fj \in FJ^{n-1}$. The n^{th} -order *Feedback Flow* $FF^n(fs, fj)$ denotes the subgraph spanned by the set of nodes $\{ fs, fj \} \cup (Desc^n(fs) \cap FAncs^n(fj))$. The $(n+1)^{\text{th}}$ -order *Feedback Flow* $FF^{n+1}(fs, fj)$ denotes the subgraph spanned by the set of nodes $\{ fs, fj \} \cup (Desc^{n+1}(fs) \cap FAncs^{n+1}(fj))$.

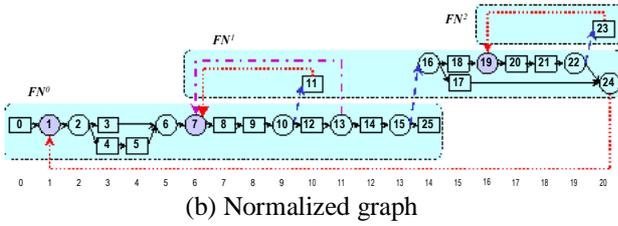
Note that we can let the main flow $MF(Start, End)$ equal to $FF^0(Start, End)$, assuming that $Start \in FS^{-1}$ and $End \in FJ^{-1}$.

When FJ is not fully classified yet, there exist at least one $FF^{n+1}(fs, fj)$ that contains node(s) of FJ . The corresponding subgraph contains some feedback structures and requires further partitioning. Our method recursively applies the steps of partition to each of these subgraphs. Otherwise, $FF^{n+1}(fs, fj)$ does not require further partitioning and equals to $FF^n(fs, fj)$, by definition.

The workflow of Fig. 2, with FN^0 , $FJ^0 = FJ$, FS^0 , and FN^1 , can be partitioned into the main flow $MF(0, 20)$, $FF^1(7, 1)$ spanned by $\{ 7, 11, 1 \}$, and $FF^1(8, 1)$ spanned by $\{ 8, 13, 1 \}$, requiring no further partitioning. For the workflow of Fig. 4, we have $MF(0, 25)$, $FF^1(10, 7)$, $FF^1(13, 7)$, and $FF^{2+}(15, 1)$. Since $FF^{2+}(15, 1)$ contains a node of FJ not classified yet, i.e., node $19 \notin FJ^0$, it contains some cyclic structures and requires further partitioning. In a similar manner, we can get $FN^2 = \{ 11, 16, 17, 18, 19, 20, 21, 22, 24 \}$, $FJ^1 = \{ 19 \}$, and $FS^1 = \{ 2 \}$ from $FF^{2+}(15, 1)$. Since nodes in FJ are fully classified, we can conclude that $FN^2 = N - FN^0 - FN^1 = \{ 23 \}$ and derive additional acyclic subgraphs of $FF^1(15, 1)$ and $FF^2(22, 19)$. Table 2 summarizes the results of the analyses for the workflow graphs of Fig. 2 and 4. Note that each finally derived subgraph of a feedback flow, except main flow, matches for distinct simple cycle in the given workflow graph.



(a) A workflow with nested cycles



(b) Normalized graph

Fig. 4. Normalization of a workflow graph with nested cycles

Table 2. Summary of partitioning

Case	Target graph	Classified nodes	Derived subgraphs
Fig. 2	$G(N, T)$	$FJ = \{1\};$ $V-FN^0 = \{11, 13\}, FJ = \{1\}, FS = \{7, 8\}.$	$MF(0, 20)$ $FF^1(7, 1)$ $FF^2(8, 1)$
Fig. 4	$G(N, T)$	$FJ = \{1, 7, 19\};$ $V-FN^0 = \{11, 16, 17, 18, 19, 20, 21, 22, 23, 24\}, FJ = \{1, 7\}, FS = \{10, 13, 15\}.$	$MF(0, 25)$ $FF^1(10, 7)$ $FF^2(13, 7)$ $FF^{1+}(15, 1)$
	$FF^{1+}(15, 1)$	$FN^1 = \{11, 16, 17, 18, 19, 20, 21, 22, 24\}, FJ = \{19\}, FS = \{22\}; FN = \{23\}.$	$FF^1(15, 1)$ $FF^2(22, 19)$

The proposed method allows a cyclic workflow model to be analyzed further, if necessary, with several smaller subflows, which are all acyclic. The complexity of this phase can be estimated as follows. Let q be the maximum degree of feedback and r be the average number of subgraphs in FN^i that need to be further partitioned, in the given workflow graph. The identification of FN^i (with FS^{n-1} and FJ^{n-1}) from $FF^{n+}(fs, fj)$, where $fs \in FS^{n-1}$ and $fj \in FJ^{n-1}$, is rather straightforward with $R_{FJT}(i, fj)$, where $i \in FF^{n+}(fs, fj)$, that has complexity $O(|T|)$, where $|T|$ is the number of transitions in the cyclic subgraph $FF^{n+}(fs, fj)$. Therefore, the complexity of this step will be $O(qr|T|) < O(|N| \cdot |T|)$.

4. TWO-PHASED VERIFICATION OF CONTROL FLOWS

This section presents two-phased verification of structural conflicts in workflow models. First, it detects certain types of structural conflicts, in the early stage, incurred from the inappropriate composition of feedback flows. And then, it applies verification to each acyclic subgraph, which is much easier to comprehend and verify individually than the whole workflow model, in general.

4.1 Early stage verification on combining partitioned flows

Lemma 1. A workflow with an AND-feedback join node fj will deadlock at fj . A workflow with an AND-feedback split node fs will potentially result in an infinite loop (i.e., livelock) or multiple instances.

Any $FF^i(fs, fj)$ split from a XOR-split fs and merged into an OR-join fj , which is exclusive with the originating

subflow in FN^i , will not introduce additional deadlock and lack of synchronization conflict to the combined flow. On the other hand, feedback joins or feedback splits of AND-type incur structural conflicts to the combined flow. Fig. 5 shows that conflicts are caused by combining a feedback flow with AND-controls. An AND-join at fj can never be started, thus causing a deadlock (Fig. 5(a)). An AND-split at fs may never end since fs always triggers a feedback, thus leading to an infinite loop without any type of proper exit from the loop, or may result in multiple instances of subflow sf_i before the synchronization, if exist, with the subflow sf_j exited from the loop (Fig. 5(b)).

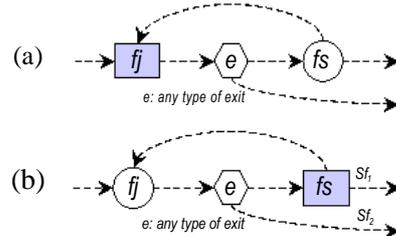
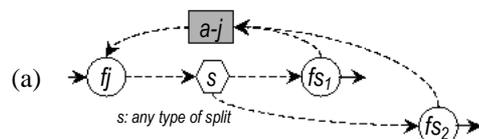


Fig. 5. Conflicts caused by AND-controls as a feedback join or a feedback split

Lemma 2. An AND-join $a-j$ that joins $FF^i(fs_1, fj)$ and $FF^i(fs_2, fj)$, with $fs_1 \neq fs_2$, incurs a potential deadlock at $a-j$ as illustrated in Fig. 6(a). In addition, an AND-split $a-s$ that splits $FF^i(fs, fj_1)$ and $FF^i(fs, fj_2)$, with $fj_1 \neq fj_2$, may result in multiple instances of subflow sf_j or deadlock at j as illustrated in Fig. 6 (b).

In Fig. 6 (a), when there exists some precedence relationship between fs_1 and fs_2 , it is clear that a deadlock conflict will occur at the AND-join node $a-j$. The only way to prevent this deadlock is when split s is of AND-type and both fs_1 and fs_2 always lead to the AND-join $a-j$ altogether, if any one of them does. But, this cannot be guaranteed and may lead to potential deadlock. In Fig. 6 (b), when there exists some precedence relationship between fj_1 and fj_2 , it is clear that lack of synchronization conflict will occur at one of them visited later. Note that subflow sf_1 or sf_2 may have some XOR-exit, not represented in the figure for simplification, which makes the analysis even more complex. The only way to prevent this lack of synchronization conflict is when the join node j that subflow sf_1 and sf_2 merge should be of AND-join and both sf_1 and sf_2 should always lead to the AND-join j altogether, if any one of them does so. But, this cannot be guaranteed either and may lead to potential deadlock. Fig. 6 (c) shows the control-flow-equivalent workflow transformed from the workflow of Fig. 6 (b), by duplicating certain nodes [28] where these additional constraints are satisfied.



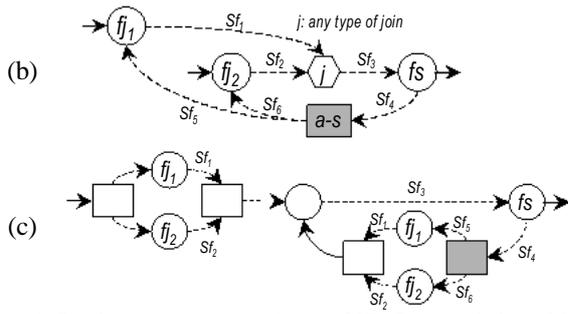


Fig. 6. Inadequate usages of (a) AND-Join and (b) AND-Split on feedback paths; (c) equivalent control flow transformed from (b) of a special case

Lemmas 1 and 2 can be used to detect certain types of structural conflict related to feedback flows. Our method identifies those ill-structured control flows based on Lemmas 1 or 2 in a workflow model. The decomposed verification phase described next, will assume that the workflow models contain only XOR- feedback split and join nodes.

4.2 Decomposed verification for each acyclic flow

We conjecture that the verification of a given workflow can be done by verifying the *main flow* and each partitioned *feedback flow* of the given workflow, where all of which are acyclic. Theorem 1 below argues the correctness of this approach formally.

Theorem 1. *If there exist no structural conflicts in $MF(Start, End)$ and each $FF^+(fs, fj)$, then there exist no structural conflict in the integrated model.*

Proof: Any possible additional structural conflicts in the integrated model caused by combining each $FF^+(fs, fj)$, originated from and merged into $MF(Start, End)$, are of lack of synchronization. After early stage verification with Lemma 1 and 2, each $FF^+(fs, fj)$ splits from an XOR feedback split fs , merges at an OR feedback join fj , and does not cross another feedback flow by an AND-split or an AND-join. Therefore, there is no chance that any additional Lack of Synchronization conflict would be caused in the combined model. The fact that each $FF^+(fs, fj)$ can be composed in similar manner completes the proof. \square

By Theorem 1, verification of structural conflicts in a cyclic workflow model can be done by verifying structural conflicts of $MF(Start, End)$ and each $FF^+(fs, fj)$, which are all acyclic. That is, in decomposed verification approach, we only need to verify *acyclic* flows. This is a very important feature of our approach and can lead to simplified verification and significant computational efficiency compared to those approaches that verify a cyclic graph as a whole. For instance, the graph reduction technique [25] can be applied to a cyclic workflow graph after the original graph is partitioned into s subflow graphs. As a result of partitioning, the computation effort can be even improved from

$O((|N|+|T|)^2 \cdot |N|^2)$ ([3], [25]) to $O((|N|+|T|)^2 \cdot |N|^2 / s^3)$ by dividing $|N|$ and $|T|$ by s , respectively.

5. CONCLUDING REMARKS

Workflow analysis and design is a fundamental task in business process management because of the recent surge of e-business process automation efforts in the corporations worldwide. Although hierarchical decomposition of a complex workflow process is an useful step in workflow analysis and design, until now there has been no reported work of its automated support in the literature. When the process model is complex, it is quite perplexing for a human to recognize substructures of potential subprocesses [2].

In this paper, we proposed a unified approach of feedback partition and decomposed verification of structural conflicts that can handle cyclic workflows efficiently. We showed how to identify the main flow and each feedback flow that completes each simple cycle, which are all acyclic. Hierarchical decomposition of a workflow model by feedback structures can make workflow analysis and design more accurate and efficient and can lead to a better design. We also showed that verification of structural conflicts in a cyclic workflow model could be done by independently verifying structural conflicts in main flow and each feedback flow, which are all acyclic. That is, in decomposed verification approach, we only need to verify *acyclic* flows.

Our work is related to three main studies in workflow verification, namely the graph reduction approach ([25], [29]), the Petri nets approach ([1], [4], [6], [33]), and the logic-based approach ([11], [27]). The graph reduction approach detects structural errors by trying to reduce the original workflow graph into an empty one. If this attempt fails, the workflow is determined to contain errors. One main drawback is that the graph reduction technique cannot handle cyclic workflows and overlapping patterns algorithmically. The Petri-nets-based workflow verification depends on the formalism of Petri nets. Therefore, for non-Petri-nets workflow models, it is difficult to apply their verification technique. The logic-based approach depends on the expressive power of propositional logic and does not take advantage of instance flows in its verification method. Furthermore, none of these approaches use the concept of feedback partitioning, which has been shown to provide advantageous features in workflow analysis and design. Our work unifies process decomposition and verification while other verification techniques focus mainly on process verification. In sum, our approach provides a useful, if not better, alternative to existing workflow verification methods.

The issues that merit future research are: to establish results for verifying correctness of a workflow model including other perspectives, for instance the data-flow;

transformation of a free-form graph-structured workflow model into a structured model, like in BPEL4WS [10]; to extend results for further workflow analysis of process reengineering.

ACKNOWLEDGEMENT

This work was supported by the 2003 Inje University research grant.

REFERENCES

- [1] Aalst, W. M. P. van der, "The Application of Petri Nets to Workflow Management", *The Journal of Circuits, Systems and Computers*, vol. 8, no. 1, pp. 21-66, 1998.
- [2] Aalst, W. M. P. van der, "Workflow Verification: Finding Control-Flow Errors using Petri-net-based Techniques", In W.M.P. van der Aalst, J. Desel, and A. Oberweis, Editors, *Business Process Management: Models, Techniques, and Empirical Studies*, volume 1806 of *Lecture Notes in Computer Science*, pp. 161-183, Springer-Verlag, Berlin, 2000.
- [3] Aalst, W. M. P. van der, "An alternative way to analyze workflow graphs," *14th Int. Conf. On Adv. Info. Sys. Eng.*, pp. 535-552, 2002.
- [4] Aalst, W. M. P. van der and A. H. M. ter Hofstede, "Verification of workflow task structures: A Petri-Net-based Approach," *Information Systems*, vol. 25, no. 1, pp. 43-69, 2000.
- [5] Aalst, W.M.P. van der, A.H.M. ter Hofstede, and M. Weske, "Business Process Management: A Survey", In W.M.P. van der Aalst, A.H.M. ter Hofstede, and M. Weske, Editors, *International Conference on Business Process Management (BPM 2003)*, volume 2678 of *Lecture Notes in Computer Science*, pp. 1-12. Springer-Verlag, Berlin, 2003.
- [6] Adam, N. R., V. Atluri, and W. Huang. "Modeling and Analysis of Workflows using Petri Nets," *Journal of Intelligent Information Systems*, vol. 10, pp. 131-158, 1998.
- [7] Aissi, S., P. Malu, and K. Srinivasan. "E-business process modeling: the next big step," *IEEE Computer*, vol. 35, no. 5, pp. 55-62, 2002.
- [8] Basu, A. and R. W. Blanning, "A formal approach to workflow analysis," *Information Systems Research*, vol. 11, no. 1, pp. 17-36, 2000.
- [9] Basu, A. and A. Kumar, "Research commentary: Workflow management issues in e-Business," *Information Systems Research*, vol. 13, no. 1, pp. 1-14, 2002.
- [10] BEA Systems, IBM Corporation, & Microsoft Corporation, Inc., *Business Process Execution Language for Web Services*, Version 1.1, 2003, available at <http://www.ibm.com/developerworks/library/ws-bpel/>.
- [11] Bi, H. H. and J. L. Zhao, "Mending the Lag between Commerce and Research: A Logic-based Workflow Verification Approach," *Computational Modeling and Problem Solving in the Networked World*, Kluwer Academic Publishers, pp. 191-212, 2003.
- [12] Business Process Management Initiative, *Business Process Modeling Notation*, Working Draft 1.0, August 2003, available at <http://www.bpmi.org/>
- [13] Casati, F., S. Ceri, B. Pernici, G. Pozzi, "Conceptual Modeling of Workflows", In M.P. Papazoglou, Editor, *Proceedings of the 14th International Object-Oriented and Entity-Relationship Modeling Conference*, volume 1021 of *Lecture Notes in Computer Science*, pages 341-354. Springer-Verlag, Berlin, 1998.
- [14] Choi, Y. and J. L. Zhao, "Matrix-based abstraction and verification of e-business processes," In *Proc. the 1st Workshop on e-Business*, pp. 154-165, 2002.
- [15] Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms* (2nd Ed.), MIT Press, 2001.
- [16] Delphi Group, *BPM2002: Market Milestone Report*, available at <http://www.delphigroup.com/>.
- [17] Fan, W. and S. Weinstein, "Specifying and reasoning about workflows with path constraints", In *Proceedings of the 5th International Computer Science Conference (ICSC'99)*, HongKong, China, volume 1749 of *Lecture Notes in Computer Science*, Springer, pp. 13-15, 1999.
- [18] Georgakopoulos, D., M. Hornick, and A. Sheth, "An overview of workflow management: from process modeling to workflow automation infrastructure", *Distributed and Parallel Databases*, vol. 3, pp.119-153, 1995.
- [19] Gondran, M. and M. Minoux, *Graphs and Algorithms*, John Wiley & Sons Ltd., 1984.
- [20] Hofstede, A. H. M. ter, M. E. Orlowska, and J. Rajapakse, "Verification Problems in Conceptual Workflow Specifications," *Data and Knowledge Engineering*, vol. 24, no. 3, pp. 239-256, 1998.
- [21] Kiepuszewski, B., *Expressiveness and Suitability of Languages for Control Flow Modelling in Workflows*, PhD thesis, Queensland University of Technology, Brisbane, Australia, 2002.
- [22] Kumar, A., and Zhao, J. L., "Dynamic Routing and Operational Controls in Workflow Management Systems," *Management Science*, vol. 45, no. 2, pp. 253-272, 1999.
- [23] Kwan, M.M. and P.R. Balasubramanian, "Adding workflow analysis techniques to the IS development toolkit," *Proc. 31st Int'l Conf. on System Sciences*, vol. 4, pp. 312-321, 1998.
- [24] Leymann, F., D. Roller, and A. Reuter, *Production Workflow: Concepts and Techniques*, Prentice-Hall, Englewood Cliffs, NJ, 2000.
- [25] Lin, H., Z. Zhao, H. Li, and Z. Chen, "A novel graph reduction algorithm to identify structural conflicts," *Proc. of the 35th Hawaii Int. Conf. On Sys. Sci. (HICSS'02)*, IEEE Computer Society Press, 2002.
- [26] Murata, T., "Petri nets: Properties, analysis, and applications", *Proceedings of the IEEE*, vol. 77, no. 4, pp. 541-580, 1989.
- [27] Mukherjee S., H. Davulcu, M. Kifer, P. Senkul, and G. Yang, "Logic Based Approaches to Workflow Modeling and Verification", In J. Chomicki, R. Meyden, G. Saake, Editors, *Logics for emerging applications of databases*, Springer-Verlag, pp. 167-202, 2004
- [28] Oulsnam, G., "Unraveling unstructured programs", *Computer Journal*, vol. 25, no. 3, pp. 379-387, 1982.
- [29] Sadiq, W. and M. E. Orlowska, "Analyzing process models using graph reduction techniques," *Information Systems*, vol. 25, no. 2, pp.117-134, 2000.
- [30] Sheth, A. P., W. M. P. van der Aalst, and I. B. Arpinar, "Processes driving the networked economy," *IEEE Concurrency*, vol. 7, no. 3, pp. 18-31, 1999.
- [31] Sivaraman, E. and M. Kamath, "On the use of Petri nets for business process modeling", *11th Annual Industrial Engineering Research Conference*, Orlando, Florida, 2002
- [32] Stohr, E. A. and J. L. Zhao, "Workflow automation: Overview and research issues," *Information Systems Frontiers*, vol. 3, no. 3, pp. 281-296, 2001.
- [33] Verbeek, H. M. W., T. Basten, and W. M. P. van der Aalst, "Diagnosing workflow processes using Woflan," *Computer Journal*, vol. 44, no. 4, pp. 246-279, 2001.
- [34] Workflow Management Coalition, *Glossary. Document Number WfMC-TC-1011*, 1999.
- [35] Workflow Management Coalition, *Interface 1: Process Definition Interchange Process Model. Document Number WfMC TC-1016-P*, 1999.

Pricing Policy for Supply Chain in the Electronic Commerce Environment

Qiong Guo, Deli Yang, Jianjun Wang, Yunhu Huo

School of Management, Dalian University of technology, Liaoning 116024, China
vipguoqiong@hotmail.com

ABSTRACT

This paper presents a model for designing the pricing policies for the supply chain whose channel is integrated with retail and e-tail. In order to incentive the retailer, the manufacture in the supply chain adopts return policy promising that any unsold product can be returned to the manufacture for a partial refund in the end of the selling season. How can the manufacture coordinate the pricing policy between the retail and e-tail channel and what is the optimal pricing policy for the wholesale price and return policy which will make the supply chain's profit maximize are the central issue in the supply chain coordination. Pricing policies between retail and e-tail are researched under the Stackelberg price competition model firstly. Under the optimal competing pricing policies model, we further obtain the equilibrium pricing for the wholesale and return price that make the supply chain maximize through the expected profit analysis. Then further research dimensions are discussed in the end of the paper.

Keywords: pricing policy, supply chain, profit, electronic commerce

1. INTRODUCTION

In the past decade, rapid developments in information and telecommunication technology have substantially changed the landscape of organizational computing. Electronic commerce (E-commerce), including Web-based commerce, mobile commerce (M-commerce) and ubiquitous commerce (U-commerce), is based not only on developments pertaining to the Internet compound, but also on prior technological and organizational innovations arising from the combination of telecommunications and organizational computing[1]. As the most rapidly developing, e-commerce via the Internet (i.e., Web-based commerce) has substantially affected the business world since the commercialization of the internet and evolution of the dot.com in the early 1990' and it is commonly believe that the transactions volume via e-commerce will continue to grow in the coming years[2]. Different companies have different uses of the e-marketplace and different e-marketplaces take different roles. There is an increasingly number of company which takes e-marketplace as a distribution way which named as e-tail. By this way, many companies, such as dangdang.com in china, seems to make success possible. However, in most case, it turn out that the most workable solution is the way that is integrated with e-tail channel and traditional retail ones. By this hybrid channel, customer can have more than one way to purchase commodity by their attitude or convenience.

With the development of e-commerce, e-tail become as a crucial retail way in distribution channel more and more in practice and there is an increasingly number of research paper that is focus on the distribution channel which is integrated with e-tail and retail. For example, Dong-Qing and John J study the equilibrium pricing policy under the e-commerce [3]. Tsen-Ming Cha studied the return policy in supply which is integrated

with e-marketplace [4]; Chen and Siems discusses the relationship between the B2B e-marketplace announcements and the shareholders response in the financial market [5]; Chiang studied the dual channel supply chain design for goods that do not provide a large service (value) such as books, CDs, etc. their results show the manufacturer can mitigate the profit loss by adding a direct channel [6]. Besides these researches, Iyer examined how manufacturers should coordinate distribution channels when two retailers compete under price and non-price attributes in supply chain earlier [7].

This paper focuses on providing an analytical pricing model for using both e-marketplace and retail as distribution channels in a supply chain instead of proposing a conceptual framework. We study a supply chain which is integrated by a return policy and the manufacture can distribute products by directly "on-line" (by e-Marketplace, referred to as e-tail) and traditional channel by which the manufacture wholesale the products to the retailer. Although, the mix of distribution with e-Marketplace has added a new dimension of competition to the firm's distribution channels, the manufacture faced with the problem how to design the pricing policy in the e-tail and regular retail channels in order to coordinate the distribution channels. Also the manufacture must decide the wholesales price and the returned price before the selling season in order to maximize mutual profit for manufacture and retailer. In fact, the analysis of return policy for channel coordination has been studied by Pasternack in the paper Optimal pricing and returns policies for perishable commodities which published in the marketing science literature in 1985 [8]. However, Pasternack's study didn't consider the impact of e-commerce on the commodities distribution because there is no e-commerce experience in that time. Cha study the optimal return policy under two undisturbed market [4]. Also, Hon-Shiang Lau studies the manufacture's pricing

policy strategy and return policy for a single-period commodities in which the retail channel didn't integrate with e-marketplace [9]. However, with the emergency of e-commerce, more and more manufacture integrates e-marketplace as the aid way to distribute their products [10], such as some in china. And the manufacture in supply chain is faced with the new problem how to coordinate the pricing policy among different distribution channels.

In this paper, we assume a two-echelon supply chain with a single manufacturer who supplies a single item to the retailer and the manufacture adopts both e-tail and regular retail channels to distribute products. Both the channels are designed for different customers respectively and the customer can acquire value-added service by regular retail channel which cannot receive from the e-tail channel; however, the price attached to e-tail would be lower. Under different types of retail channels and service level, what are the equilibrium pricing policies that manufacture should adopt for the e-tail and retail channels in order to coordinate the sales quantity and the profit in different channels? We will examine the pricing equilibriums between the e-tail channel and retail channel under the types of competitive pricing schemes: Stackelberg competition models. Under the Stackelberg model, the manufacture, as the leader of the market, announce the e-tail price before the selling season firstly. Also, the manufacture has considered the retailer's reaction prior to announce the e-tail price. As the follower in the market, the retailer reacts to the e-tail price by choosing retail price and the corresponding service level to maximize its own selling quantities and the profit.

As a result, we obtain the equilibrium prices between e-tail and retail channels. Because of promising in contracts, the retailer can return any unsold products to the manufacture for a partial re-fund products after the selling seasons which can incentive the retailer to order more quantities products and work harder for retailing, by this way, the manufacture shares the risk caused by demand uncertainly with the retailer by this way. Also, the retailer can acquire reasonable profit by different order quantity and the retail price. After we study the competing pricing policy between the e-tail and retail channels under the Stackelberg competing model before the selling season, the equilibrium pricing policy for wholesale price and return price is studied in the paper which can coordinate the profit and risk between the manufacture and retailer.

On the basis of the discussion, the paper is organized as follows: the quantity model that the supply chain is integrated with retail and e-tail channel is set up in section 2, and the Stackelberg competing pricing model is presented in successions; as a result, we obtain optimal pricing policy pertain to e-tail and retail channels under the discussion on the model. Furthermore, we derive the expected profits for all the parties in the supply chain

under the wholesales price and returns policy in section2. After deducing, the optimal pricing policy for the wholesale price and returns policy is proposed in the same section. In the end, concluding remarks are presented.

2. MODEL

Based on the above discussion, we consider a supply chain inside which a single-product manufacturer-retailer system mixed with an e-tail distribution channel in addition to a regular retail channel where the manufacture sells the product to a retailer at a fixed market with a unit wholesale price of w_0 and the unit production cost at the manufacture is c . The retailer then sells the product to the customers in the lower echelon with price p_2 . In the past, the manufacture can only distribute items by this way. However, with the development of e-commerce, the manufacture can select an e-tail channel to sell the products directly on-line at the same time in which the price is p_1 . At the same time, the retailer competes against the on-line selling by adding value service via the traditional channel when distribute the product to lower echelon at a retail price p_2 . The cost which is produced by value-added service defined as $c(y)$, such as speciously packing for books and CDs or properly maintain service for cars which are needed for some customers who are assumed to be heterogeneous and are able to purchase a product either from a retail channel or directly from an e-tail channel and some of the them would like to purchase directly from the manufacturer at a lower cost, while others are willing to pay more for the value added by the retailer that are unavailable in e-tail, on the contrary. Because of the value-added service, the retail price p_2 is typically expected to be higher than e-tail price p_1 .

After the selling season the retailer can return any unsold products to the manufacture under the returns policy on the price w_1 which is promised in the contracts, so the manufacture shares the risk caused by demand uncertainly with the retailer by this way. As we discussed above, the retailer can acquire reasonable profit by different order quantity and the retail price p_2 either. In this paper, we assumed that the manufacture can sell any returned products by the e-marketplace faced with global and local markets and whose demand is assumed to be immensity. In the past, without the e-commerce, the manufacture would salvage the returned products at a price μ that more lower than p_1 . However, any unsold products cost the retailer a unit holding cost h . In the past, there is an example of using both e-tail and traditional retail in china, such as eguo.com.

In this paper, we assume all information is common to both parties which mean that the retailer has complete information about manufacture (such as unit production

cost, etc) and vice versa. With the above details, Fig 1 shows the structure of the supply chain with the use of the e-tail.

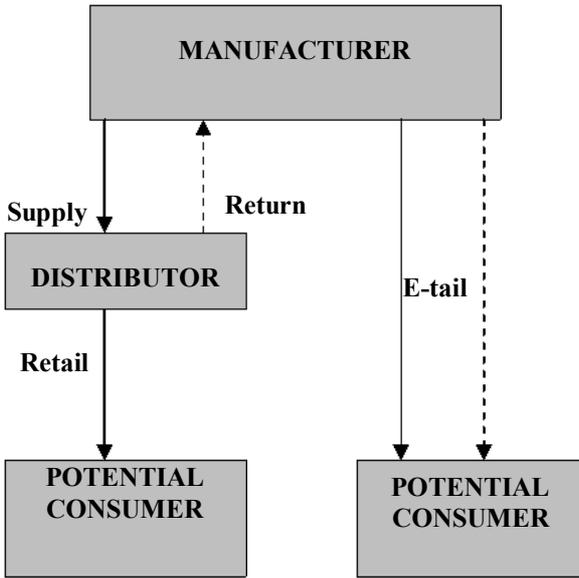


Fig.1. a supply chain mixed with e-tail and retail

Firstly, we have the following notation

- π_1 Profit of the manufacture;
- π_2 Profit of the retailer;
- π Profit of the supply chain;
- d_1 Demand via e-tail channel;
- d_2 Demand via retail channel;
- p_1 Unite price of the e-tail channel;
- p_2 Unite price of the retail channel;
- w_0 Unite wholesale price of the product;
- w_1 Unite return price for the unsold product
- v Cost of unit value-added service
- q Quantity of product retail ordered from manufacture;

Under the Stackelberg model, the manufacture, as the market leader, firstly announced the unit e-tail price p_1 to maximize payoff π_1 on consideration the expect reaction price before the selling season; as the follower of the market, the retailer then reacts to the manufacture's price policy p_1 by adopting unit price p_2 and value-added service which unite cost is v to maximize its market share and own profit π_2 .

So we can define the profit for manufacture, retailer and the supply chain as following:

$$\pi_1(p_1, p_2) = p_1 D_1(p_1, p_2, v) - w_0 D_1(p_1, p_2, v) - w_1 D_2(p_1, p_2, v) \quad (1)$$

$$\pi_2(p_1, p_2) = p_2 D_2(p_1, p_2, v) - w_1 D_2(p_1, p_2, v) \quad (2)$$

$$\begin{aligned} \pi(p_1, p_2) &= \pi_1 + \pi_2 \\ &= p_1 D_1(p_1, p_2, v) + (p_2 - c(v)) D_2(p_1, p_2, v) \end{aligned} \quad (3)$$

We define the e-tail channel demands d_1 and retail demands d_2 as following which are similar to that in Tsay etal [11] and Choi [12]:

$$d_1 = D_1(p_1, p_2, v) = (a_1 - b p_1) + \lambda(p_2 - v - p_1) \quad (4)$$

$$d_2 = D_2(p_1, p_2, v) = (a_2 - b p_2) + \gamma v + \lambda(p_1 + v - p_2) \quad (5)$$

Where a_i is the market base through channel i ($i = 1$ for e-tail and $i = 2$ for retail), b_i represents the marginal channel demand per respective channel price, and v measures the marginal retail demand per retail service value added. The parameter λ , termed discussion intensity, is adopted to describe the shift between the two channels with regards to the price and value.

We assume that before the selling seasons both the retail and the manufacture expect that the entire product q will be sold by either e-tail or retail, which will make their profit be maximized. Under this consideration, we can formulate the Stackelberg pricing equilibrium of p_1 and p_2 . In the competing pricing model, we assumed that after the manufacture announced the e-tail price p_1 the best-reaction price that the retailer would adopt to maximize its profit is p_2 which make the following equation satisfy:

$$p_2(p_1) = \arg \max_{p_2} \pi_2(p_1, p_2) \quad (6)$$

Anticipating against price p_2 that would be adopted by the retailer, the manufacture will select the best e-tail price p_1 , which is given as:

$$\max_{p_1} \pi_1(p_1, p_2(p_1)) = p_1 d_1 + w d_2 \quad (7)$$

Taking the derivative of (3) with respect to p_1 and let $\partial \pi_1 / \partial p_1 = 0$ yields:

$$\begin{aligned} p_1 &= \frac{2a_1 b + 2a_1 \lambda + a_2 \lambda + \gamma v \lambda + c(v) b \lambda}{4(b + \lambda)^2 - 2\lambda^2} \\ &+ \frac{c(v) \lambda^2 + 2w b \lambda + 2w \lambda^2 - v \lambda^2 - w v b \lambda}{4(b + \lambda)^2 - 2\lambda^2} \end{aligned} \quad (8)$$

As the same way, we can obtain:

$$p_2 = \frac{\lambda p_1 + a_2 + (\lambda \gamma + v) + (c(v) + w)(b + \lambda)}{2(b + \lambda)} \quad (9)$$

p_1 And p_2 is the competing equilibrium price that the manufacture and retailer decide by Stackelberg model before the selling seasons, which would make their profit maximize. Under the equilibrium pricing policy, they will contract the wholesale price w_0 and return price w_1 in order to share profit and risk in the cooperation process, So we define the π_1, π_2, π as the real profit

for manufacture, retail and supply chain end the selling season. Then we discuss the optimal wholesale price w_0 and return price w_1 in the following.

End the selling season, profit for the manufacture, retailer and supply chain are defined as following:

$$\pi_r = (p_2 - w_0)q - (p_2 + h + c)(q - L) \quad (10)$$

$$\pi_2 = (p_2 - w_0)q - (p_2 + h + c)(q - L) \quad (11)$$

$$\pi_r = \pi_2 + \pi_1 = (p_2 - w_0)q - (p_2 + h + c)(q - L) \quad (12)$$

Where L shown the amount of returned items after the retail selling season and it is defined as:

$$L = \max(0, q - x) \quad (13)$$

From the equation (10), (11) and (12), we can formulate the expect profit for the manufacture, retailer and supply chain as:

$$E[\pi_1] = (p_2 - w_0)q - (p_2 + h + c)(q - L) \quad (14)$$

$$E[\pi_2] = (p_2 - w_0)q - (p_2 + h + c)(q - L) \quad (15)$$

$$E[\pi_r] = (p_2 - w_0)q - (p_2 + h + c)(q - L) \quad (16)$$

The expected returned products shown as following:

$$E[L] = \int_0^q (q - x)f(x) dx \quad (17)$$

Taking the derivative of (15) (16) with respect to q , we can show the result of the optimal product quantities for retailer and supply chain as:

$$\frac{dE[\pi_2]}{dq} = (p_2 - w_0) - (p_2 + h + c)(1 - F(q)) \quad (18)$$

$$\frac{dE[\pi_r]}{dq} = (p_2 - c) - (p_2 + h + c)(1 - F(q)) \quad (19)$$

Let (18) (19) equal to zero, we can acquire the optimal product quantities respectively for retailer and supply chain. It is obviously that there is only one solution for the optimal quantities in the supply chain. So we can obtain equation (20) from equation (18) (19) which are shown as:

$$\left(\frac{p_2 - c}{p_2 + h + c} \right) w_1 - w_0 = \left((p_2 + h) \left(\frac{p_2 - c}{p_2 + h + c} \right) - p_2 \right) \quad (20)$$

The manufacture and the retailer have decided price p_1 and p_2 before the selling season, which we have explained by Stackelberg competing model shown in (8) and (9). From the equation (20), we can find that the wholesale price w_0 and return price w_1 are the variable which can the manufacture use as the way to share the profit and risk with retailer by different wholesale price w_0 and return price w_1 . In order to make the profit of the supply chain maximize, price w_0 and w_1 must satisfy the equation (20).

3. CONCLUSION

This paper shows how a monopolistic manufacture of a single-period commodity could establish his pricing policies in a mixed e-tail and retail distribution channel in the supply chain under the e-commerce environment, in which the manufacture and retailer consider not only the wholesale price and return price, but also the competitive equilibrium pricing policy between the e-tail and retail firstly which we demonstrate by Stackelberg competition model. Also we obtained dynamic equilibrium equation between the optimal wholesale price and return price, which can make the supply chain's profit maximize and serve as a way for the manufacture and the retailer sharing the profit and the market risk.

In order to simplifier, we assumed that all returned product would be sold entirely by e-marketplace in this paper. However, as an extension, we can research the condition in which the returned products are still left in the e-tail in the end, which will make the research more complex. Also, we can also discuss the manufacture and the retailer announces the price simultaneity, so the price equilibrium can not be received by Stackelberg competition model. In more generally, we may further study equilibrium prices for different demand forms contrary to the linear demand which is assumed in this paper. Further study could be done how the incentive plan can be set to coordinate the manufacture and the retailer to make more contribution to the supply chain.

ACKNOWLEDGEMENT

The authors are grateful for the anonymous referees. Thanks are especially given to the organizer of The Fourth International Conference on Electronic Business (ICEB2004) for their hard work. This paper is supported by the National Natural Science Foundation of China (No: 70031020)

REFERENCES

[1] V.zwass, Electronic commerce and organizational innovation: aspects and opportunities, Int. J. Electron. Comm. Vol. 7. No.3, pp7-37, 2003.
 [2] Ward, M.R. On forecasting the demand for e-commerce, Working Paper, University of Illinois, Urbana, Champaign. Pp15-19, 2000.
 [3] Dong-Qing Yao, Competitive pricing of mixed retail and e-tail distribution channels, The International Journal of Management Science Vol. 04, pp1-13. 2004.
 [4] Tsen-Ming Cha, Optimal return policies for supply chain with e-marketplace, Int. J. Production Economics Vol. 88, pp205-227, 2004.
 [5] Chen, A.H., Siems, T. F., B2B e-marketplace announcements and shareholder wealth, Economic and Financial Review. Vol. 9, pp 12-20, 2001.
 [6] Chiang WY, Chhajed D, Hess JD. Direct marketing, indirect profits: a strategic analysis of dual-channel

supply chain design. *Management Science* Vol. 49. No. 1, pp1-20, 2003

[7] Iyer G. Coordinating channels under price and service competition. *Marketing Science* Vol. 17, pp338-355, 1998

[8] Pasternack, B.A. Optimal pricing and returns policies for perishable commodities. *Marketing Science* Vol. 4, pp166-176, 1985.

[9] Hon-Shiang Lau. , Manufacture's pricing strategy and return policy for a single-period commodity. *European Journal of Operational Research* Vol. 116 pp291-304, 1997.

[10] Jen-Her WU, Tzyh-Lih Hisa. Analysis of E-commerce innovation and impact: a hypercube model. *Electronic Commerce Research and Application* Vol. 5, pp1-16, 2004.

[11] Tsay Andy A, Agrawal N. Channel dynamics under price and service competition. *Manufacturing and Service Operations Management* Vol. 2, pp372-391, 2000.

[12] Choi SC. Price competition in a duopoly common retailer channel. *Journal of Retailing* Vol. 72. No. 2, pp117-134, 1996.

Recomposing of Industrial Supply Chain and Constructing of Digital CBD

Baixin Yang, Deli Yang, Yunfu Huo

School of Management, Dalian University of Technology, Dalian, Liaoning 116024, China
somdyang@dlu.edu.cn

ABSTRACT

Drawing lessons from the development of cities at home and abroad success and lost experience, through practice research, this paper systemically analysis the evolution of urban economy space. In the scope of global industrial supply chain's recomposing, it is discusses that modern city's function structure characteristic, focusing "integrative logistics management service". According to Dalian's geography and resource advantage in district industry supply chain, this paper also proposes the digital CBD planning of Dalian which focuses integrative logistics management service.

Keywords: industrial supply chain, recomposing, integrative logistics service, digital CBD

1. EVOLUTION OF URBAN ECONOMY SPACE AND RECOMPOSING OF INDUSTRIAL SUPPLY CHAIN

The city is a zone that has high population density and depends on the second and third industries. And as the center of around area, it performs advanced social, economic, and cultural activities and has many organizations that benefit from complicated aims. Thus the city plays important roles in the development of district economy: district manufacture center, district currency center, district service management center, district human resource center, district intercommunion center, and district innovation center⁽¹⁾.

It is proved through investigation that the forming and development of city lies on four factors: relative benefit, intramural scale economy, traffic cost and convergence economy. With the change of ages and the development of society, those factors have different weights, which determines the forming, flourishing and declining of the city.

As the theory of evolution of urban space, the forming and development of modern city basically follows the "equation- convergence- separation-new and advanced equation" evolution course. Because of scale economy and scope economy, the space convergence of some leading industries is higher (the trade cost is lower) in certain districts (cities), which makes many industries concentrate in certain kernel enterprises/industries, and forms the convergence effect of urban space. At the same time, with the increasing of convergence scale, the scarcity of recourses (soil and so on) makes the opportunity cost increase and the diseconomy phenomenon appear, and the advancement of technology makes the structure of scope economy change greatly. In the dynamic "convergence-separation" process of urban development, the scale of city is increasing, and population flow accordingly direct contact⁽²⁾.

In the analysis of industry system, Dicken's supply

chain theory is more representative. He think that in the producing process of certain manufacture or service, supply chain is the combination of a series of increasing value and interactive and interrelated basic activities, consisting of gathering of raw and processed materials, transportation, storage, gradual process, assembling and distribution of manufacture and service and other functions. In supply chain, every activity relates each other through different technology devotion and traffic and communication, enclosing the economic system that can supply essential management fund. And moreover, all activities should been adjusted, standardized, and controlled⁽³⁾. The industrial creating value process is represented by the value chain that is same to supply chain.

In the traditional economy increasing theory, the economy increasing mainly dues to the actual elements such as capital, soil, labor and so on, and namely the basic activities of supply chain. After 1960s, western nations' economical increasing have obviously changed, in which the status of traditional manufacturing descended, on the other hand, information industry that was represented by finance and information management, rapidly developed, and more and more accelerated the development of GDP. Information industry became the important factor of the development; service economy (producer services) gradually replaced industry economy; besides the basic activities of supply chain structure, elements took more and more weight in industrial increment. Having undergone the phases of resource denseness, labor denseness and capital and technology denseness, western industrialized countries have been the phase of technology and intelligence which is called as "knowledge-based economy" and in which not only science and technology, though the "accumulation-overflow", but also the devotion of capital, labor and other essentials can gain increasing income, so can the hole economy system⁽⁴⁾.

Firstly, this industrial variety (upgrade) represented as the alter of industrial innovation mode, that is to say,

innovation was the continuous and linear process of invention, examination, manufacture and distribution, but now it change the discontinuous and nonlinear process. In the studying economy and soft industry system that the most important resource is knowledge, the innovation always exists in the manufacture process, “studying in practice”. Given as an example, the “modularization” manufacture based on decentralization and centralization and recomposing knowledge just is an innovation of the manufacture mode and the consumption mode⁽⁵⁾.

According as the space benefit and the accessibility that are gained by the manufacturing space benefit and cost, we can separate the space structure into three types: central cluster, decentralized cluster and random spread. It has been proved that whichever type of recomposing supply chain has the “direct separation-pervasion-second convergence” characteristics⁽⁶⁾:

Firstly, the supply chain changed from the lengthwise unification that takes products as core and enterprises as cell to the supply chain one that takes cooperating management as core and relative industries as cell, and the logistics was separated from the manufacture and distribution⁽⁷⁾.

Secondly, considering that the urban center has higher rent and more expensive accessibility, only the factors of the supply chain that have the highest boundary benefit can be gathered. But the urban center district has the advantages in cooperation management, public service, information communion, sharing knowledge, and manpower resource, so industries always locate management center and distribution center on this district.

Thirdly, according to the cost counting, with the descending of rent, the manufacture center of industries diffuses from urban center district to qui-center district. Generally, core enterprise locates its manufacture base on the urban qui-center district and the accessional enterprise encloses the core one.

Fourthly, the substance of soft manufacture is “Just-In-Time (JIT)” mode, which demands that the reaction between producer and consumers is in time, and thus the logistics activities must focus on the manufacture and distribution. The development of traffic technologies makes the transportation cost generally lower and the increasing of urban rent makes the storage cost more and more higher. Accordingly, physical distribution closes up the manufacture center and distribution center, services the manufacture and distribution of goods, and the logistics center or the distribution center is located around or near the urban center.

Fifthly, the final aim of recomposing supply chain is to apply the “integer supply chain—through specialized

dividing work, achieving the increasing of scale benefit”, which is represented by the cooperation management of supply chain elements, what is called “cooperation manufacture”⁽⁸⁾. In the steps of supply chain, the logistics needs more information and higher supply level of the public property, so information management and cooperation activities give the shape of the center of logistics service center, which mostly is located at urban center districts.

2. SPACE CHARACTERISTICS AND FUNCTIONS OF MODERN URBAN CBD

The modern urban CBD (Central Business District), generally is the urban center district. It refers to the most convenient joint, and commonly is recognized by people, and is relatively certain.

With the development of technology and the advancement of global economy, CBD has more and more been the center of complicated economy activities and the support of economy increasing, in big cities and even bigger districts. Because of the standing “convergence- separation” action, supply chain’s elements lie on the top of the value chain, mostly concentrate in CBD. So the situation affects exclusive economic effect, communion effect, and space effect, which enhances the convergence and takes bigger appeal. Therefore, the urban CBD is not only the symbol of glory and development, but also the support of extending influence. As an urban space modality, CBD has following characteristics:

Firstly, the district economy center. Geographically, CBD is generally on the center of a city, and economically, it is the most convenient spot. The frequency of communion that CBD provides determines its role in the city and more abroad zone.

Secondly, the third industries intensity. The third industries are gathered at CBD, and the relative enterprises are dense. Generally, the intensity index of the third industries at the CBD is hardly lower to II, that is to say, the whole area of the third industries is more than 40% of the CBD’s area.

Thirdly, high convenience and strong crowding. The outer and inner traffic of CBD is three-dimensional, and the traffic area is commonly 20% of the whole area. CBD has the most developed inner and outer traffic, and can form a convenient traffic net. Which offer people the good opportunity to pass. Relatively, the crowing degree is the highest in the city.

Fourthly, using of the space is developing vertically, and has high modern and intelligence level. Most buildings are high and even super-high. Using the advanced structural material, equipment, and communication products, the constructions of CBD is intelligent, so they can satisfy the require of modern official business.

Fifthly, rent is expensive. The area of CBD is scarce and the house cost is expensive.

Recomposing supply chain results in the change of industrial elements' structure. Accordingly, the space layout and the function of cities has altered⁹.

Firstly, manufacturing spreads to the urban qui-center district. Before the middle phase of industrialization, most office buildings, storages, factories, and shops are lied at CBD. But in the last industrialization ages, most manufacturers begin to move to the qui-center districts of the city.

Secondly, the unmanned trend. That mainly refers to the population density of urban center is lower and lower, and the house area is reducing.

Thirdly, the advanced development of the service industry. Service economy weights more in the value chain. The advantage of the CBD in service industry makes it competitive in the service industry and it can function the big scope and scale management of supply chain.

Fourthly, the structure of the material services adjusts. The type, service object, scale and quality of the material services at CBD have great changes. The internationalization and intelligent are current. The business hotels are the main force of restraints and traditional hotels, and super-markets are abundant.

Fifthly, commonality goods' importance is obvious. Because the CBD functions the economy cooperation and management, social economy security is more and more important. The government, tipstaff, chamber of commerce, and other functional departments have higher efficiency than those in other districts.

Those changes directly result in the reorientation of urban CBD. Because the core of industry benefit has transferred from physical elements to intelligent elements, from manufacture to management, the functions of urban CBD are restricted by the "cost-benefit" curve, and transfer from the manufacture center to the management center.

The industrial increment generally consists of four parts: value in shape, value on time, value in space, and value on possess. It is obvious that the industry management center and distribution center base the "integration of supply chain" theory. According to modern logistics conception, the logistics is a industrial system; is the integration and optimization of transport, storage, loading and unloading, conveying, packing, current processing, distribution, information disposing and other elements. Society economy activity is an integration of material distribution, information current, manpower resource, and capital. From the scope, the economic function of urban CBD is represented by

management of information, manpower resource and economic services.

3. ANALYSIS OF THE EVOLUTION OF DALIAN'S CBD

Such as the former depiction, most western cities takes CBD as the general management, distribution, and service center, which many business, economy, office activities. According to foreign CBD, our native urban CBD has developed through different path. Native CBD developed on the base of traditional center district. Traditionally, the center of a city usually is the geometrical center or the axes of the urban area. In the latter-days, with the invading of western capitalism and the slow development of national economy, the rudiment of CBD appeared in the seashore cities. At the same time, the retail commerce took shape in traditional commerce and handicraft industry areas. In the planned economy ages, the structural elements and layout of urban CBD had obvious changes, and effected by Soviet model, CBD's function is the center of administration management and public campaign. Since the Reform, the inner impetus of market economy promoted our native city internationalization development¹⁰. But the general service function is still lack. The reasons are following:

Firstly, the reconstructing of old city zone mostly is physical reform and compensation. It can hardly aim to renew and develop the city function, and cannot practice multi-objective, multi-factor and multi-approach reform. Secondly, the imbalance of our native industries development behaves as the lag of the construction of public physical and economic services. With the rapid development of native economy and the distinctness of that China becomes the manufacture center of the world, Chinese economy development has involved in the global situation¹¹. It has been proved by investigation that our nation can not afford the demand of recomposing supply chain in the logistics management services¹².

Thirdly, the low density of CBD became the inanition. The development of urban service function is slower than the demand of society economy developing, so manufacturing move out of the CBD, at the same time, the supply of service industry is not enough. And the government's compulsive policy make it worse. Fourthly, the bad circle of physical market of CBD formed. The lag of general services makes the layout and constructing of many public business buildings straggled.

The history of Dalian city is relatively short, and the programming and constructing not only has seashore structure characteristic, but also has been influenced by the colonial age when Dalian was colonized by Russia and Japan. After almost one hundred years constructing, a CBD distribution has formed in Dalian that centers on

the “Qinniwa bridge-Zhongshan square-Bay square” triangular district. Since 1990, with the rebuilding of old city zone, enclosing the former CBD and aiming at achieving the international level, Dalian government has weightily programmed and constructed a 2.8 sq.km. area from Bay square to Civil aviation mansion. And obviously a functional distribution has basically formed, which emporiums, office buildings and hotels .

In general, the layout and functional characteristics of Dalian CDB, is same to the problems of other cities. But Dalian CDB has its characteristics:

Firstly, the competition of retail trade is too serious. Many big bazaars have higher sale but lower benefit. The situation is worse in the super-markets such as Wal-Mart.

Secondly, the restaurants hardly analysis the clients. Service has no otherness and the service level is relatively lower.

Thirdly, the competition of general services is too weak. The scale economy and scope economy cannot form. It is difficult that achieving higher services to supply the demand of clients. Those are the reasons why many office buildings leave unused.

Fourthly, the lag of public physical services is serious. Because the local benefits bring the public resources waste and the trade charge is too high, the public resources cannot be shared.

Modern economy is the typical “seashore” economy, and the port is the hinge and core of an economic chain. With the global development of the world economy, markets have managed elements and not managed products. The development of nations more and more lies in the outer trade and international exchange. This trend directly accelerates the “tertiary port” forming. This sort of port combines the advantages of container transport, multi-type transport, international trade, general logistics management.

The “tertiary port” not only should have traditional services, such as loading and unloading goods, distribution of information, providing services for the clients, but also should supply many modern services:

Firstly, providing industrial services.

Secondly, providing distribution services.

Thirdly, providing general logistics services of supply chain.

In the above services, the last one is the most important, whose level directly influences the using of port^[3]. Port industry is one of the important industries of logistics. The “tertiary port” has been the core element of modern logistics, which takes the global supply chain management as aim and emphasizes on the general logistics management.

As the estimate of World Bank, the last year, the logistics cost of China National Economy is 1,600 billion RMB, and the logistics of China has been shaped. According to the statistics, in 2001, the total amount of commodity is 3,760 billion RMB in China, that of yield material is 5,500 billion RMB, and that of imports and exports is 500 billion dollars. The bigger overseas demand provides wide marketable space.

As the hinge port of northeast Asia, Dalian is one of the important cities that are taken as national logistics basic constructions, and Dalian takes the domain role in the district industry supply chain. The advantage of Dalian is just its ports, likes Singapore and Hong Kong. So the development of Dalian city lies on the interplay of port economy and city function^[4].

As one leading industry, the general services industry should aim at the “port economy”. Recent years, the depression of many hotels and office buildings that are primary elements of general services industry can prove this point.

After China entered WTO, with the flourish of native and international trade and the gradual opening of native market, Dalian logistics industry has advanced opportunity and great potential. Dalian has taken logistics as support industry and put forward the stratagem that constructing Dalian port as international ship center^[5].

But in the layout and programming of district logistics, mainly emphasizing on “hard” devotion, and less considering the “soft” devotion, which may waste much capital and substance. The development of Dalian logistics should take the district advantage as bases, and take the demand of global industry supply chain as criterion. This logistics system has following characteristics:

ò Space: shorter traffic distance and better traffic conditions

ò Time: shortcut and efficiency

ò Circumstances: comprehensive living and business establishments and grace sights

ò Commerce: comprehensive administration and information services, reducing cost and increasing efficiency

Not only from district resources advantage and the demand of marketing circumstances, but also perfecting the competition and driving the whole district economy, it is necessitous that constructing the comprehensive logistics services stage to strengthen Dalian CBD’s functions.

It is obvious that at present, rare services are the effective supply, regulation and conformity following functions:

Firstly, public physical services

Secondly, comprehensive society logistics management

services

Thirdly, logistics clients, physical logistics recourses

4. CONCLUSIONS AND SIGNIFICATION OF RESEARCH

Constructing digital CBD, can strengthen the services functions of comprehensive logistics management of Dalian, and absolutely combine the district advantage and resource advantage of Dalian. So Dalian urban CBD's functions can rapidly adapt the press of global economic, and the CBD becomes the competitive center of city or the area and accelerates the recomposing of district industry supply chain.

ACKNOWLEDGEMENT

This paper is supported by the National Natural Science Foundation of China (No: 70031020).

REFERENCES

- [1] Arthur O'Sullivan, Urban Economics, McGraw-Hill, 2000.
- [2] Michael E. Porter, The Competitive Advantage of Nations, The Free Press, New York, 1990.
- [3] Denise Dipasquale, Urban Economics and Real Estate Markets, Beijing: Economics Science Press, 2002.
- [4] Li, X., Economics Geography, Beijing: High Educational Press, 1999.
- [5] Michael E. Porter, Advantage of Competition, Beijing: Huaxia Press, 1997.
- [6] E. Brain Arthur, Management OF Industry OF High Technology, Beijing: Chinese People's University Press, 2002.
- [7] Wang, J., Innovational Space: Assembling of Corporation and Regional Development, Beijing: Peking University Press, 2001.
- [8] Masahiko Aoki, Introduction of Modularization Theory, *Comparing*, Vol.: 2, 2002.
- [9] Ma, S., Supply Chain Management, Beijing: Machine Industry Press, 2000.
- [10] Yang, X., Economics Theory, Beijing: Chinese Social Science Press, 1998.
- [11] Paul, K. W., 21st Century Manufacturing, Prentice-Hall, 2001.
- [12] Ballou, R., H., Corporation Logistics Management, Beijing: Machine Industry Press, 2000.
- [13] China Logistics and Purchase League, Investigation of MNC's Logistics Demands in China, Beijing, 2002.
- [14] Dalian Economics Research Center, Dalian Modernism Logistics Development Plan, Dalian, 2002.
- [15] Cheng, J., Digital City: Theory, Method and Application, Beijing: Science Press, 2003.

Research on the Supply Chain Partnership Relations with Association Rule Analysis

Huiying Zhang¹, Dezhi Zhu¹, Jin Du², Xin Chen²

¹ School of management, Tianjin University, Tianjin 300072, China

² Tianjin Enterprises Survey Organization, Tianjin 300050, China
zhanghuiying@nankai.edu.cn

ABSTRACT

Due to that many fuzzy and uncertain association relations represented by some phenomenon in the process of enterprises cooperation in supply chain, data mining method was try out in this paper. At first, concept and deducing procedure was introduced, and then 300 samples from a survey was pretreatment, in succession cooperation relations of enterprises in supply chain was studied combined with association analysis method. Some acceptable conclusions were obtained, which can explain the relations between modality and measures of cooperation of enterprises reasonably.

Keywords: Association rule, Supply chain, Partnership

1. INTRODUCTION

Supply chain is a whole function net model that connects from supplier, manufacturer, distributor, retailer and the consumer. Purchase the material, manufacture the work-in-process and end product, and then transfer the product to the consumer by the distributed net, with the controlling flow of information, logistic and finance and surrounding the core enterprise [1]. Nearly every enterprise exists and develops in the supply chain. Many scholars, enterprisers and the consulers are researching the supply chain partnership. In this paper, we focus on the enterprise in Chinese Tianjin and sample of 300 enterprises. We studied the partnership from supplier and client at two aspects, one is how they exchange information and the other is how they solve the trade dispute. For the one aspect, we research the relationship of the cooperation range and time between information exchanges. For the other, we studied the relationship of the quantity of cooperater which has trade dispute and whether they sign the contract between the methods of solving the trade dispute. This research reflects the phenomena of Tianjin because basing on the data from the surveys, so it provides reference to the enterprises.

2. SUPPLY CHAIN PARTNERSHIP

Supply chain Partnership is the collaborative agreements of sharing information and profit, proportioning risk between enterprises. Supply chain can reduce transaction cost because of information exchange, mutual trust and fine partnership [2]. The effective partnership established by the collaborative enterprises will become the special resource and knowledge that can't be copied by others, so it is a private competitive advantage and protect the counterpart become opportunist. It's very important for all enterprise, not only manufacture but also service, to have stabile partnership. Establishing the useful partnership is the core competence for the enterprises to

enhance the satisfied level of the consumer and to achieve the win-win purpose between them by adding the product value [5].

3. DATA MINING AND ASSOCIATION ANALYSIS METHOD

3.1 Data Mining

Data mining is an integration research field across many subjects from the late middle of 90th. It includes database system, statistics, machine learning, visual and information science [4]. Its purpose is to find out the valuable knowledge and information from huge and complex data, by analyzing the data with machine learning and statistics and storing and managing the data with database and data warehouse. The statistics is old and has exact theory, includes described statistics and concluded statistics. But the machine learning method has fewer limits and provides more comprehended result than statistics [5].

3.2 Association Analysis Method

All methods of data mining are different and adapted different field [5]. Using the association analysis can find out the association rules in the data. These rules give the condition of the appearance of attribute-value in the dataset. More exactly, association rule describes the impact of the appearance of one item on the others, i.e. the interesting relation between items. It is popular in the marketing research and web analysis [5] [6]. There are also some scholars mining the text of the open question in survey.

The association rule can be formally stated as follows:

- 1) Let $I = \{i_1, i_2, \dots, i_m\}$ a set of literals, called items.
- 2) Let D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$.

3) Associated with each transaction is an unique identifier, called its TID.4) Let X is a set of items in I . A transaction T is said to contain X , if $X \subseteq T$.

An association rule is an implication of the form $X \Rightarrow Y$, where $X \subset I, Y \subset I$ and $X \cap Y \neq \Phi$. We must consider three value to be sure whether an association rule is strong:

- 1) Support s : probability that the transaction set D contains $X \cup Y$, support $(X \Rightarrow Y) = P(X \cup Y)$.
- 2) Confidence, c : conditional probability that the transaction set D having X also contains Y , confidence $(X \Rightarrow Y) = P(Y | X)$.
- 3) Lift [9]: the ratio of confidence to expected confidence of Y .

$$cor_{x,y} = \frac{P(Y | X)}{P(Y)} \quad (1)$$

The rule is strong with the at least minimum support (called minsup) and minimum confidence (called minconf) and the lift is more than 1, where minsup and minconf are user-specified values.

Except for all above, a really interesting rule must have two conditions [9]:

- 1) It is exceptional and unusual.
- 2) It is potential useful.

These tow conditions must be decided by the expert in the field, so the people must involve in the process of mining association rule.

Basing on the kind of the variable, the association rules can be divided to the qualitative and the quantitative. Basing on the kind of dimension in the data, they can be divided to the single-dimension the multi-dimension. In the survey, the qualitative and quantitative variable is the same, and a item correspond to a dimension, so the data is multi-dimension.

3.3 The Applicability of The Association Analysis Method for The Partnership

Because we use the data from survey, the reflected status of the enterprises is limited. We only use limited several items because the survey length. The relationship of the cooperation range and time between information exchanges is not casual, and the relationship of the quantity of cooperator which has trade dispute and whether they sign the contract between the methods of solving the trade dispute is not either. We just master the objective conditions. Even though there are some causal relationships between the items, these are not sufficient. We don't emphasize the

inevitable law but want express the rational correlation, so the association analysis method is suitable.

4. DATA DESCRIPTION

We gather data of the relation from the enterprise to the supplier or the client by the survey. The items are about the district and the contact method and how the enterprise solving dispute. These are:

- 1) The question "How the client (supplier) is distributed?" includes four answers, "in the same section", "in the same city but not the same section", "in the same country but not the same city" and "abroad". It expresses the distributing of the client and the supplier.
- 2) The question "How long the major client (supplier) has business with you?" includes five answers, "within 1 year", "from 1 year to 2 years", "from 2 years to 3 years", "from 3 years to 4 years" and "beyond 4 years". The time is longer, the relation is better.
- 2) The question "What's the method of how you connect with your client (supplier)?" includes five answers, "telephone or fax", "interview", "internet", "post" and "others". It's the information exchange method.
- 3) The question "What's the number of your major client (supplier)?" is a quantitative variable.
- 4) The question "What's the number of your client (supplier) dispute with you?" is a quantitative variable.
- 5) The question "Do you sign contract with your client (supplier)?" is a binary variable.
- 6) The question "Which method is to solve dispute?" include "go to court", "court of arbitration" and "negotiation".

There are 200 manufacture enterprises and 100 service enterprises in our sample which include big, middle and small company and different kinds.

5. PRE-PROCESS AND ANALYSIS

We deal with the data use the EM (Enterprise Miner) tool of the SAS (Statistic Analysis System) software after defining the question and gathering the survey data.

5.1 Data Pre-process

The original data is usual not designed special for the data mining, so we must revise or transform the data to the acceptable form for the mining tools. These are

clearing, integration, transforming and reducing. In this paper, we use the transforming method.

- 1) Because the original data is in the Excel table, we use the "import" function of SAS to import them in the SAS database first. There are 300 cases in the sample. Each case is a row and each item is a column in the database.
- 2) We use the ratio of the number of having dispute to the number of major client (supplier) to express whether the disputing partners are much. It's a quantitative variable. For the new qualitative variable, we must make the quantitative variable discrete before association analysis. In this paper we use the cluster to provide two new variables basing on the space partition. One is for the client number and the other is for the supplier number.
- 3) The question "How long the major client (supplier) has business with you?" has five answers. Considering the internet is widely used after 1997, so we transform to two variables using the dividing line, 1997.
- 4) The association analysis in the EM tool need transaction database, so we design a program in SAS to convert the relational database to transaction database.

5.2 Mining Analysis Rules In EM

The EM mining tool in SAS is a graphical operation, so it's easy to use. We just indicate the database, the variables and the analysis method. In the association analysis, the input variables are ID (transaction identify) and Item (transaction item). We identify the minsup is 5% and the minconf is 30% and the lift more than 1, so we get lot of rules. We filter the useful from all by controlling the antecedent set and the consequence.

5.3 Result

A perfect supply chain is involved the foundation of the collaboration between enterprise, supplier and client about the logistics and information exchange, and the foundation of the contract rule [4]. So we analyses method of information exchange and solving the trade dispute.

5.3.1 The association analysis for the information exchange

First, we analyses the relation between the tow items, distributing and collaboration time, to the information exchange method. Because the positive and correlative rules are interesting, so we only discuss the rules with big lift. The results are:

- 1) The companies with client on abroad and collaborated beyond 3 years mostly use the internet for information exchange. Support is 5.33%. Confidence is 34.04%. Lift is 2.55.
- 2) The companies with client on abroad have more relation with the internet information exchange method. Support is 6.67%. Confidence is 30.30%. Lift is 2.27.
- 3) The companies with domestic client have more relation with the traditional information exchange methods, "telephone and fax" and "interview". Though the lift is less, the confidence is much.
- 4) The companies with supplier on abroad and collaborated beyond 3 years mostly use the internet for information exchange. Support is 4.39%. Confidence is 34.21%. Lift is 3.89.
- 5) The companies with supplier on abroad have more relation with the internet information exchange method. Support is 5.41%. Confidence is 33.33%. Lift is 3.79.
- 6) The companies with domestic supplier have more relation with the traditional information exchange methods, "telephone and fax" and "interview". Though the lift is less, the confidence is much.

The result show that more and more companies use internet in the supply chain to contact the partner on abroad so they reduce the cost by the e-business. But the companies contact the domestic partner just use the traditional method, "telephone and fax" and "interview".

5.3.2 The association analysis for how they solve the trade dispute

We only consider the cases have trade dispute because most cases in the sample are no dispute.

First, we analysis the relation between the quantity of cooperator which has trade dispute and whether they sign the contract to the methods of solving the trade dispute. We only consider the rules with big lift. The results are:

- 1) The companies with more disputed client have more relation with the "negotiation" method. Support is 21.88%. Confidence is 100%. Lift is 1.42.
- 2) The companies with less disputed client have more relation with the "go to court" method. Support is 21.88%. Confidence is 37.84%. Lift is 1.27.
- 3) The companies with more disputed supplier have more relation with the "negotiation" method.

Support is 15.63%. Confidence is 100%. Lift is 2.21.

- 4) The companies with less disputed supplier have more relation with the “go to court” method. Support is 10.94%. Confidence is 46.67%. Lift is 3.32.
- 5) The companies with less disputed supplier have more relation with the “negotiation” method. Support is 28.13%. Confidence is 94.74%. Lift is 2.09.
- 6) There is no significant relation between whether they sign the contract to the methods of solving the trade dispute.

The result shows that the companies solve the dispute by tight collaboration and negotiation when they have more disputed client or supplier considering the long profit, and the companies with less disputed client or supplier tend to go to court to solve the occasional accident.

5. CONCLUSIONS

In this paper, we try to analyse the survey by the data mining and don't limit to the traditional statistics for the uncertain problem. We analyze the survey data from 300 samples of enterprise in Chinese Tianjin and find out a series of appropriate rules about the supply chain partnership. These rules can explain the relations between modality and measures of cooperation of enterprises reasonably and reflect the status of Chinese supply chain partnership and help the manager to make decision. So this method is suitable to the association

REFERENCES

- [1] Ma, Shihua, etc. *Supply Chain Management*, Mechanism Industry Press, 2000.
- [2] Xu, Shujun, etc. “A study on transaction cost of supply chain enterprises”, *Industrial Engineering and Management*, Vol.6, No. 6, pp25-27, 2001.
- [3] Ma, Xinan, etc. “Supply Chain Partnership and Partner Selection”, *Industrial Engineering and Management*, No. 4, pp33-36, 2000.
- [4] Jiawei Han, Micheline Kamber. *Concept and Technology of data mining*. Mechanism Industry Press, Aug, 2001.
- [5] Indranil Bose, Radha K. Mahapatra, “Business data mining — a machine learning perspective”, *Information Management*, pp211-225, 2001
- [6] Qi-Yuan Lin, Yen-Liang Chen, Jiah-Shing Chen, Yu-Chen Chen, “Mining inter-organizational retailing knowledge for an alliance formed by competitive firms”, *Information Management*, pp431-442, 2003.
- [7] Dai Wensheng, etc. “Association Rule in data mining”. *Statistical Research*, Vol.20, No.8, pp40-42, 2002.
- [8] Tian Jinlan, Zhang Suqin, Huang Gang. “Investment risk rules mining in insurance business data with association rules method”, *J Tsinghua Univ (Sci & Tech)*, Vol. 41, No. 1, pp 45-48, 2001.
- [9] Cai Weijie, etc. “Survey of Association Rule Generation”. *Computer Engineering*, Vol.27, No.5, pp31-33, 2001
- [10] Zhao Lindu, “Supply chain management in the era of knowledge economics”. *Journal Of Southeast University (Natural Science Edition)*, Vol.32, No.3, pp514-522, 2002

Studies on Horizontal Competition among Homogenous Retailers Based on Agent-based Approach

Ming Xie, J. Chen

Research Center for Contemporary Management, Tsinghua University
School of Economics and Management, Tsinghua University
Beijing 100084, China
{xiem1, chenj}@em.tsinghua.edu.cn

ABSTRACT

This paper adopts agent-based simulation to study the horizontal competition among homogenous price-setting retailers in a one-to-many supply chain (one supplier and multiple retailers). We model the supplier and retailers as agents, and design their behavioral rules respectively. The results show that although the agents learn individually based on their own experiences, the system shows an asymptotic convergence, which approaches Nash equilibrium. Based on these results, we discuss the effects of the retailers' horizontal competition on the retail price, retailers' and supplier's profits.

Keywords: supply chain, agent-based simulation, horizontal competition, Nash equilibrium

1. INTRODUCTION

The concept of supply chain was introduced several years ago and has been widely studied in academic researches and industrial applications. A supply chain is a network composed of suppliers, manufacturers, retailers, etc., which cooperate to offer a kind of goods or service. The partners in a supply chain cooperate to gain extra profits and also compete for more profits. These cooperations and competitions, along with the various uncertainties, make the research on supply chain a quite complex and difficult task. [1] shows that in the academic study of supply chain, only some relatively simple issues are discussed, and many assumptions and constraints are imposed to these models.

Besides analytical methods, simulation is also an important research methodology. Compared with analytical methods, it is more suitable to study complex issues. Simulation has been widely applied in various engineering fields, and recently some social scientists have tried it in social science research and have got valuable results. The most prominent research in this direction is experimental economics, e.g., [2], [3] and [4].

There are also researches that use simulation to study management issues. For example, discrete event simulation [5] has a relatively long history in operations research. But such simulations are focused on the sequence and/or causal relationships of events, which is not suited to describe the competition and cooperation relationships between the partners in a supply chain. On the contrary, agent-based modeling adopts a bottom-up approach and focuses on the design of agents' individual behaviors, and watches what systemic macro behaviors evolve from individuals' micro dynamics [6]. This is quite suitable for modeling complex systems, such as supply chain systems [7].

This paper adopts agent-based simulation to study horizontal competitions in supply chains. Horizontal competition is a hot spot and also a nodus in supply chain research. It is generally analyzed with non-cooperative game theory. But just as we have mentioned before, the application of such an analytical approach is limited into relatively simple models. [8] studies a supply chain model with multiple retailers, each of whom faces a newsboy decision problem. Based on this model, [9] adopts discrete choice model to describe the dynamic process of arrival and choice behavior of each customer. As in [8], the retail prices are fixed, and the retailers only determine order quantity. We know that in many cases retail price affects customer demand. So the model in [9] should be extended to a more sophisticated model with multiple competing price-setting retailers, i.e., retailers can adjust their retail prices to influence customer demands. But this model is almost intractable with analytical methods. In this model, each retailer's profit is determined not only by his own retail price and order quantity, but also on other retailers'. The competition becomes more intense, and things become more interesting: how will these price-setting retailers make their decisions? Will the retailers find appropriate retail price and order quantity? How will the retailers' profits and the supplier's profits be? In this paper, we study such a model and try to answer these questions through agent-based simulation. We model each retailer and the supplier into an agent and design learning rules for it. The simulation results show that although the agents learn individually based on their own experiences, each of them finally gets a relatively stable policy, and the whole system asymptotically converges to an absorbing state which approaches the Nash equilibrium. Based on these results, we also discuss how the retail price and supplier/retailer profits are affected by the competitions between the retailers.

The rest of this paper is organized as follows: Section 2 describes the model settings; Section 3 shows the simulation results and provides detailed analysis; and Section 4 concludes this paper.

2. MODEL SETTINGS

The basic structure of the model is shown in Figure 1.

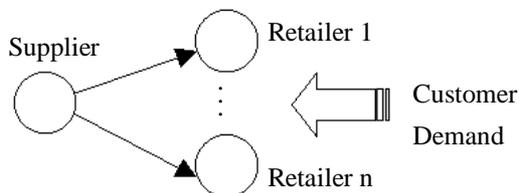


Figure 1 Model Structure

This model runs for many periods. In each period, retailers buy products from the supplier, and then sold them to customers. The supplier charges a fixed wholesale price of 1, and has infinite capacity. Every period, each retailer decides his retail price and order quantity. The products are perishable; all the unsold products are disposed with salvage value 0, and all the unsatisfied customer demands are lost. From analytical view, in each period each retailer faces a price-setting newsboy model. The only cost of each retailer is purchasing cost, occurs when receiving products from the supplier. Each retailer obtains revenue by selling products to customers. The sequence of events of this model is as follows: on each period, each retailer places an order to the supplier simultaneously; then the supplier makes enough products and sends products to each retailer with the quantity that the retailer has ordered; then customers arrive, choose retailers and buy products.

2.1 Customer Behavior

In this model, we adopt the discrete choice model [10] to represent customers' arrival and retailer-selecting behaviors. We assume that in each period, there is a population of customers, whose number is a Poisson random variable with mean 1000. The arrival of the customers is a Poisson process, and the service time of each customer is zero. Each customer desires only one product. When each customer arrives, he chooses a retailer from the series of retailers $N = \{1, 2, \dots, N\}$ to purchase one product. Each customer associates a utility U_j with each retailer $j \in N$. In additions, there is a no selection option, denoted $j = 0$, with associated utility U_0 , which means the utility of not purchase at all. A customer chooses the retailer with the highest utility among the series of retailers and buys a product from him. For example, if $Max U_j = U_2$, the customer chooses the second retailer. But if $Max U_j = U_0$, the customer returns home without buying anything. If a customer is not satisfied, i.e., if the retailer is out of stock, the customer makes a choice again.

The utility U_j is decomposed into two parts: one part, denoted u_j , represents the nominal (expected) utility; the other part, denoted ξ_j , is a zero-mean random variable representing the difference between U_j and u_j . Thus, $U_j = u_j + \xi_j$. Nominal utility u_j is determined by some identified factors. In this paper, we define $u_j = \beta a_j - \gamma p_j$, $u_0 = \beta a_0$, where a_j is a quality index and p_j is the price for product j . β and γ represent the sensitivities of the customer to the factors. The noise part ξ_j represents the effects of those unidentified factors, and is often modeled as a Gumbel random variable with distribution

$$P(\xi_j \leq x) = \exp(-\exp(-x/\mu - \tau))$$

with mean zero and variance $\mu^2 \pi^2 / 6$ (τ is the Euler's constant). Here, it is obvious that the retailer with higher retail price will have less utility for a customer, and so will be less likely to be chosen.

Different with classical utility maximization, in this model we assume U_j to be unobservable a priori, so that an individual's choice is uncertain. When a customer arrives, he makes a choice based only on knowledge of the public information of each retailer's price. We also assume all the customers are homogenous, and $\beta = \gamma = 1$, $a_j = 7.06$, $a_0 = 4$, $\mu = 0.847$ (the data comes from the model in [9]).

2.1 Retailer Behavior

We model each of the retailers as an adaptive agent. Here, we first assume that each retailer has an aspiration for more profits. No matter what policy, if it can bring more profits, the retailer will adopt it, without considering its social effects.

We assume that the retailers have minimal information on the environment: they do not know each other agent's policies, payoffs and costs. All that one retailer knows is its own experiences. Also a retailer can observe other agents' historical prices, yet we assume he does not know to utilize this information under such a complex environment. Each retailer continuously learns from his own experiences, and calibrates his policies to obtain as more profits as possible. The basic idea of each agent's learning mechanism is that: each agent has an infinite set of actions. At each stage, he chooses the action that it assesses to have the highest payoff. This idea of learning has a standard implementation – the Q-learning in Machine Learning [11]. In our model, each retailer has two decision variables: retail price and order quantity. Although these two variables are generally all continuous, we assume only discrete values can be chosen. We define the minimal difference between two possible prices to be 0.1, i.e., the prices should be 2.3, 3.4, etc. We also limit retail price into the range (1, 6) (It can be tested that when price ≥ 6 , customers will seldom buy a product; if the price ≤ 1 , the retailer will get non-positive profit, which is not rational.). Thus each agent has 50 possible retail prices.

Similarly, we assume the minimal batch size of the retailers' orders is 5. We limit the order quantity into the range [0, 2000]. Then each retailer has 400 order quantity choices.

We further model each of the price and quantity decision separately into a decision "black box", which uses Q-learning algorithm to adaptively find best choices. These decision black boxes are also learning automata in artificial intelligence [12]. When a learning automaton adopts Q-learning algorithm, it keeps a Q value for each action, such that $Q_i(a)$ represents the expected reward that the automata i believes it will obtain by playing action a^i . In this paper, we use the following formula to update the Q value of each action:

$$Q^i(a_{t+1}^i) = Q^i(a^i) + \alpha (r^i(a_1^i, \dots, a_t^i) - Q^i(a^i)) \quad (1)$$

Here, a_t^i means the action that the agent i has chosen at time t . For the actions that the agent did not choose, the Q value does not change. $r^i(a_1^i, \dots, a_t^i)$ is the payoff the agent gets after time t , α is step size. Each automaton adopts softmax policy to choose actions based on their Q values. At beginning, each action's initial Q value is set randomly. The readers can refer to [11] for a detailed discussion on Q-learning algorithm.

Here we will make a further assumption on the decision behavior of each agent: each agent believes that it faces an environment with rather high uncertainty; it does not know when the game will be over, and it knows that it is hard to make predictions. So it does not take into account the possible future earnings. Thus the agents are assumed to be myopic, and repeatedly play a multi-person game in which all players choose actions simultaneously in each round.

2.3 Supplier Behavior

We model the supplier as a non-adaptive agent. In each period, she only does routine works: receives retailer's orders, produces enough products and ships them to the retailers.

2.4 Experiment Design

We carry out the simulation on the platform of Swarm developed by Santa Fe Institute. We use the number of retailers to represent the intensity of the retailers' competition. We select 5 representative retailer numbers 1, 2, 5, 10, and 20 to carry out the simulation one by one. For each retailer number, we run the simulation for 5 times. Each time the simulation is run for 8,000 periods to allow asymptotic behavior to emerge if it is present, and each time we use different initial action weights of each agent and different random seeds. Due to the size limitation of this paper, we will not present the program code.

3. SIMULATION RESULTS

We discuss the simulation results in two steps: convergence and insights.

3.1 Convergence

Figure 2, 3, 4, 5, 6 presents the time series of the retailers' moving average profits (time span is 20) when the numbers of retailers are 1, 2, 5, 10, and 20. The horizontal axis represents time, which is denoted with the number of time grains (There are only three numbers, 20000, 40000 and 60000. They are the time represented by time grain number. In this simulation, one period consists 8 time grains) and the vertical axis represents retailers' moving average profits.

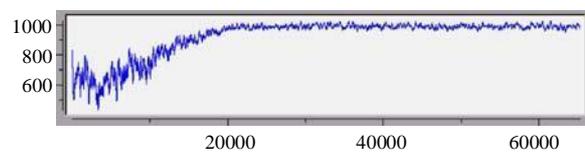


Figure 2 Time series moving average profits with 1 player

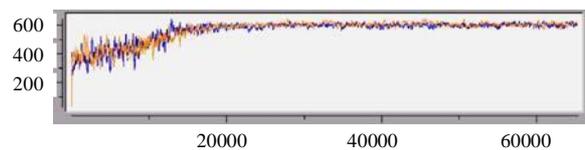


Figure 3 Time series moving average profits with 2 players

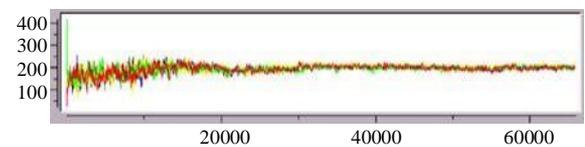


Figure 4 Time series moving average profits with 5 players

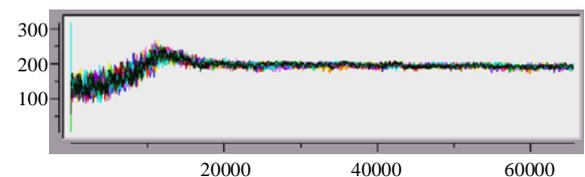


Figure 5 Time series moving average profits with 10 players

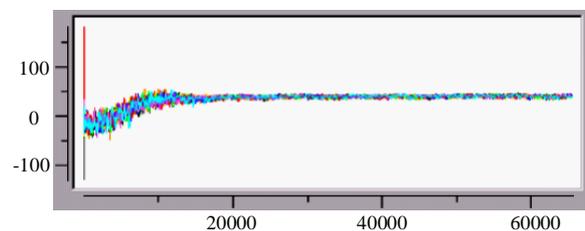


Figure 6 Time series moving average profits with 20 players

We can see from the result data that each run shows asymptotical convergence. When retailer number is the same, different run converges to the same steady state and each retailer finally finds the same policy. For example, in the 2-retailer case, each retailer's final price has mean 2.63 with standard derivation 0.05; final order quantity has mean 399 with standard derivation 18. This

result seems self-evident due to the homogeneity of the retailers. But in fact, each retailer has its own individual decision boxes. It is quite possible that different retailer converges to different steady states, or even there is no convergence at all. In this model, each agent's initial state is not identical (each action's initial Q value is set randomly), and each retailer's learning process is also not identical (this is shown in Figure 7, which is an amplification of Figure 3's initial part). So the result of convergences reflects an intrinsic property of the systems.

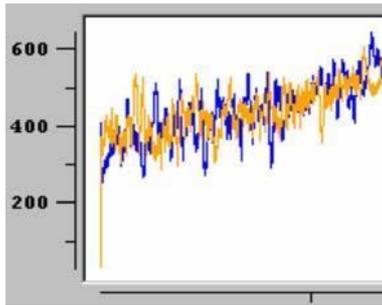


Figure 7 Initial learning dynamics in 2-retailer case

This intrinsic property is that the systems have Nash equilibria, which can be approximated by these steady states.

In this model, each agent has two learning automata, one for retail price and the other for order quantity. We implement the learning mechanism of each learning automaton using Q-learning combined with softmax selection. According to [11], a learning automaton with this learning mechanism approaches best response strategy when the temperature parameter (a parameter in softmax) approaches zero. At steady states, each learning automaton converges to stationary strategies, and the temperature parameter becomes small, so each learning automaton approaches its optimal response. According to theorems on Nash equilibria, if a game with multiple players choosing their best responses converges, it converges to Nash equilibria. Then we know that all the learning automata are in Nash equilibria at steady states.

Each agent consists of two learning automata which have identical payoffs (because these two learning automata update their Q values based on the same agent's payoff value, see (1)). Then the two learning automata of each agent also play a cooperative game with the same aim of making the agent better. So, at steady states all the learning automata approaches Nash equilibrium means two things: first, each agent sticks to stationary price and quantity policies; second, each agent finds its best combination of price policy and quantity policy given other agents' stationary policies. In other words, at steady states, no agent has incentives to change its stationary policy, i.e., the agents in the game approaches Nash equilibria.

We can conduct "turbulence tests" to see whether the equilibrium is stable. Suppose when the system arrives at a steady state, one retailer suddenly loses confidence on his stationary policies, and wants to explore more efficient ones. This is implemented by changing the temperature parameter of softmax selection. We carry out a turbulence test on a 2-retailer equilibrium, and the result shows strong stability (see Figure 8).

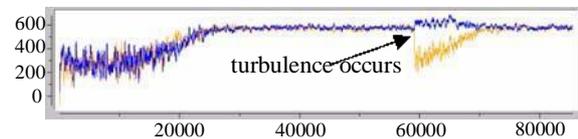


Figure 8 Turbulence test on 2-retailer equilibrium

It is notable that during the learning process, each retailer's average profit is not always increasing. See Figure 5, there is a "peak" where all retailers' profits are higher than in equilibrium. It seems strange that the agents do not learn best strategies. This phenomenon is common in multi-agent learning. Consider the Prisoners' Dilemma, each player gets highest payoff when they cooperate, but it is not Nash equilibrium: each of them has incentive to betray given the other one cooperate. The "peak" in Figure 5 stays in the similar situation: although each retailer gets rather high profits, he has incentive to change his policy to get even higher profits. All the retailers are myopic, change policies for next period's more profits, but they are finally all worse off and approach the Nash equilibrium.

The processes of how the system arrives at steady states are different in different simulation runs. Figure 9 is another run for 10-retailer simulation, which has no prominent "peaks" as in Figure 5. But although the processes of convergence are different, the final states are always the same.

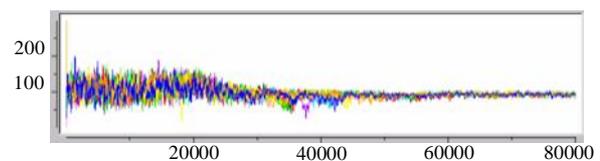


Figure 9 Another run for 10-retailer simulation

The fact that the system approaches Nash equilibrium does not mean that we are designing particular agents to find Nash equilibria. All the behavioral assumptions of the agents in this model are abstractions of real supply chain partners. We use learning algorithms to imitate the decision process of a supply chain partner who faces complex environments and with less information. The convergence means that in this model retailers will adopt stationary policies after a long period of learning and competition. This offers us a basis to investigate how the retailer price and retailer/ supplier profits are affected by the competition among the retailers: we will discuss these issues at steady states.

3.2 Insights

First we investigate how retailer competition affects retail prices. We get the retail prices at steady states of each simulation run, and list the data in table 1 and Figure 10 (These data and all the data in latter tables and figures are the average of 5 simulation runs, see section 2.4).

Table 1 Retail price under different retailer numbers

Retailer Number	1	2	5	10	20
Retail Price	2.85	2.63	2.14	2.01	1.89

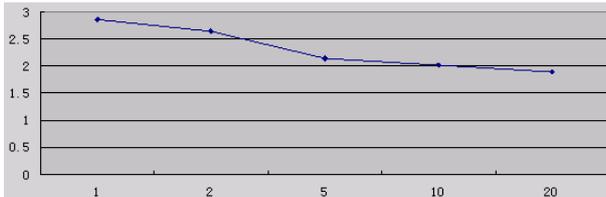


Figure 10 Retail price under different retailer numbers

It can be seen that as retailer number increases (the competition becomes more intense), the retail price decreases. So customer benefits from retailers' competition. This is correspondent with intuition. Yet we observe that although retail price decreases, it does not get too low: when there are 20 retailers, retail price is still 1.88. In other words, the price war is quite mild, and no fierce lowering-price activity appears. This is because of the assumptions on agent behaviors: aspiration on more profits and myopic. Such agents will not adopt policies that decrease short-time profits to defeat other agents.

Table 2 and Figure 11 show supplier revenue under different retailer numbers.

Table 2 Supplier revenue under different retailer numbers

Retailer Number	1	2	5	10	20
Supplier Revenue	578	800	985	1023	1044

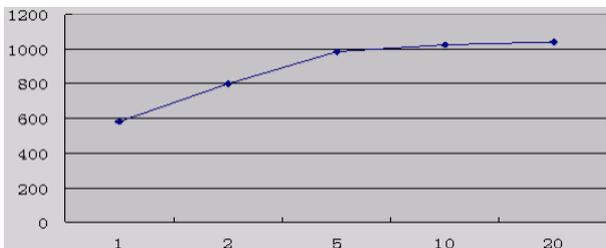


Figure 11 Supplier revenue under different retailer numbers

The supplier's revenue increases with the number of retailers, i.e., competition among the agents are beneficial to the supplier. The reason is as follows: retailer competition decreases retail price, so customer demand increases; accordingly, retailers will order more products from the supplier, and supplier's wholesale price is fixed, thus supplier gets more revenue.

We use following Table 3 and Figure 12 to study how competition affects the retailers.

Table 3 Retailer profit under different retailer numbers

Retailer Number	1	2	5	10	20
Total Profit	991	1193	1010	915	806
Single Profit	991	596	202	92	40

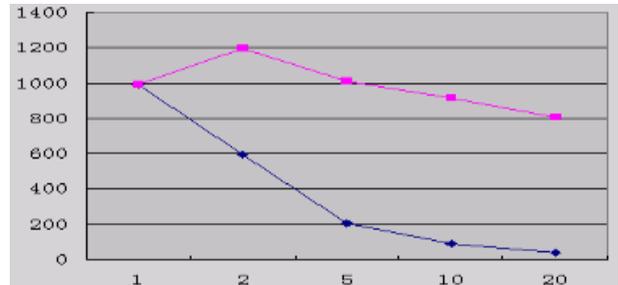


Figure 12 Retailer profit under different retailer numbers

Each retailer's profit decreases with the number of retailers in the market. In this model, the mean of total customer demand does not change and all the retailers share this market. The more the retailers, the smaller each market share.

Figure 12 also shows that retailers' total profit is maximal when there are two retailers and it decreases when there are more than two retailers. The decrease of retailers' total profit can be attributed to the competition among the retailers. Figure 10 show that retailer competition can decrease retail price. This has two effects: (1) decrease retailer profit directly, and (2) increase customer demand and thus increase retailer profit. Notice that in this model, each customer has the choice of not buying. Thus if there are more retailers, a customer is less likely to leave without buying anything, i.e., retailers' total profit increases. This can be viewed as the benefits of offering customers more selection options. We denote this effect as (3). General literature on supply chain management ignores this effect, yet following we will see that this effect can play an important role in certain cases. Whether retailers' total profit increases or decreases depends on the trade off between effect (1) (decreases retailers' total profit) and effect (2) plus (3) (increase retailers' total profit). From Figure 12 we can see when there are more than 2 retailers, the retailers' total profit decreases, which implies that effect (1) dominates effect (2) plus (3). Yet retailers' total profit is more in 2-retailer case than in 1-retailer case. This is different with general economic conclusions which believe a monopoly often obtains maximum market profits. Effect (3) plays an important role in the difference between this model and classical economic model, and makes effect (2) plus (3) dominate in this case.

4. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

In this paper, we investigate the horizontal competition in a one-supplier-multiple-retailer supply chain. We adopt agent-based simulation, which is quite different with classical methods. Based on the assumptions of the agents' behaviors, we endow each agent with a reasonable learning mechanism. The results show that the system converges asymptotically to steady states that approach Nash equilibria. Under these steady states, we analyze how retailer price and supplier/retailer profits are affected by the competition among the price-setting retailers, and notice the important role of customer choice behavior which is generally ignored in supply chain research.

This paper also inspires us an in-depth research on the issues of horizontal cooperation among the retailers. From former analysis we know that retailers' total profit decreases with retailer number. This gives the retailers strong incentives to collaborate with each other. We will address this issue in future work.

Furthermore, this paper inspires us to study vertical competition and cooperation in the supply chain. Here we get the conclusion that the supplier will benefit from the competition between the retailers. How things would be if the supplier can adjust wholesale price? And, when the supplier knows that retailer competition is beneficial for her, she will have incentive to prevent the collaborations among the retailers. So we can investigate what measures the supplier can take to prevent such collaborations.

ACKNOWLEDGEMENTS

This work is partly supported by the National Science Foundation of China (70231010, 70321001).

REFERENCES

- [1] Cachon, G.P., *Supply Chain Coordination with Contracts*, in *Handbooks in Operations Research and Management Science: Supply Chain Management*, S. Graves and T.d. Kok, Editors. 2002, North-Holland.
- [2] Kirman, A.P. and N.J. Vriend, *Evolving market structure: An ACE model of price dispersion and loyalty*. *Journal of Economic Dynamics & Control*, 2001. **25**: p. 459-502.
- [3] Chen, S.-H. and C.-H. Yeh, *Evolving Traders and the Business School with Genetic Programming: A New Architecture of the Agent-Based Artificial Stock Market*. *Journal of Economic Dynamics & Control*, 2001. **25**: p. 363-393.
- [4] Tesfatsion, L., *Agent-Based Computational Economics: Growing Economies from the Bottom Up*. *Artificial Life*, 2002. **8**: p. 55-82.
- [5] Banks, J., et al., *Discrete-Event System Simulation (3rd Edition)*. 2000: Prentice Hall.
- [6] Parunak, H.V.D., R. Savit, and R.L. Riolo. *Agent-Based Modeling vs. Equation-Based Modeling: A Case Study and Users' Guide*. in *Proceedings of Multi-agent systems and Agent-based Simulation (MABS'98)*. 1998.
- [7] Chen, J. and M. Xie. *Research on a Systematic Approach for Building Multi-Agent Supply Chain Models*. in *The 2nd International Workshop on Meta-synthesis and Complex System*. 2002. Shanghai, China.
- [8] Lippman, S. and K. McCardle, *The Competitive Newsboy*. *Operations Research*, 1997. **45**: p. 54-65.
- [9] Mahajan, S. and G.v. Ryzin, *Inventory Competition under Dynamic Consumer Choice*. *Operations Research*, 2001. **49**(5): p. 646-657.
- [10] Train, K., *Discrete Choice Methods with Simulation*. 2003: Cambridge University Press.
- [11] Sutton, R.S. and A.G. Barto, *Reinforcement Learning: An Introduction*. 1998: MIT Press, Cambridge, MA.
- [12] Narendra, K.S. and M.A.L. Thathachar, *Learning Automata: An Introduction*. 1989, NJ: Prentice-Hall.
- [13] Leslie, D.S. and E.J. Collins, *INDIVIDUAL Q-LEARNING IN NORMAL FORM GAMES*. Unpublished, 2004.

Study of Information Supply Chain and Artificial Neural Network's Related Application

Qian Li, Yongxian Wang, Youqin Zhu

School of Economics and Management, Tsinghua University, Beijing 100084, China
{Liq1, Wangyx}@em.tsinghua.edu.cn, Zyq@mail.dzdaily.com.cn

ABSTRACT

Users become less and less patient with huge useless data today. One of the great challenges now most net searching engines meet is how to get valuable information from lots of data sets. Aiming to satisfy every user's special demand, we need to integrate and optimize the whole course of data searching, including adjusting the users' input keywords, searching original results from network, and further processing of these results. Learning from the idea of Supply Chain Management, we put forward the concept of Information Supply Chain (ISC) in this paper to generalize the course above. For ISC's optimization, artificial neural network is chosen as a tool to find out the relationships between different keywords and paper categories, which are summarized and stored in knowledge base. Based on it, the process of selecting proper keywords and searching news information could be more efficient. A pruning method named MW-OBS is illustrated to train ANN as well. Some details about the framework and components are also mentioned, especially on how an individual step in ISC works, what's the relationship between them, and how they coordinate to meet every personal demand. ISC, an integrated information processing in the interests of users' individual need, has great advantages over simple searching from network with original keywords.

Keywords: Information Supply Chain (ISC); Supply Chain Management (SCM); Artificial Neural Network (ANN)

1. INTRODUCTION[□]

Now days, it's difficult not to get news any more, but to find out the really needed information from so many data resources. People use searching engines to help themselves, however lots of time is wasted in distinguishing which piece is useful or not. The purpose of our study is to improve the information searching efficiency by some ways of optimization.

Some ideas about Supply Chain Management (SCM) inspire us. In productive industry, marketing rather than producing, is very important to earn profits for factories. The products who could meet the consumers' needs well will occupy an important position in related market. More and more factories put their focus on consumers' satisfaction, according to which, all business segments are recombined as well, including purchasing, producing, inventory, and selling. From 'push' to 'pull', factories assembled their working process through SCM, and this kind of optimization bring them great profits in the end.

We are wondering why not optimize the information searching process on the similar idea of SCM? Accordingly the conception of Information Supply Chain (ISC) is put forward, and more details will be discussed in the next parts.

2. FRAMEWORK OF INFORMATION SUPPLY CHAIN

2.1 How It Works Now?

[□]This work was supported by Natural Science Foundation of Shandong Province of China under Grant No.03BS002.

When use some kind of internet searching engine, as we always do, select some keywords casually, run the engine, then get hundreds even thousands of related results. This process could be summarized in the following chart.



About this chart, some points should be mentioned. The quality of information searching is decided by keywords, which however are chosen by users themselves without any rules. In other words, these keywords usually couldn't represent crucial characteristic attribute of the information needed. The direct impact of the results is that more useless data is got meanwhile it's even harder to find out the real needed ones.

2.2 "Searching Efficiency"

In order to give a pacific description of "searching efficiency", some indexes should be defined, such as 'precision' and 'repetition'

$$precision = \frac{| \{Relevant\} \cap \{Retrieved\} |}{| \{Retrieved\} |} \quad (1)$$

$$repetition = \frac{| \{Repeated\} |}{| \{Retrieved\} |} \quad (2)$$

Where "Relevant" means the quantity of the information that meets the user's demand; "Retrieved" denotes the amount of actually searching results.

The former index illustrated how much partition is useful

in all results. The later is put forward by us as a complement. It's meaningless for users to get the duplicated information, and more time is also wasted. "Repetition" is set to explain how brief and concise of the searching results.

2.3 The Framework of ISC

Putting more attentions to users' information demand, and assembling the related segments in processing, ISC comes into being. The segments and their interrelationships are depicted through the below chart.

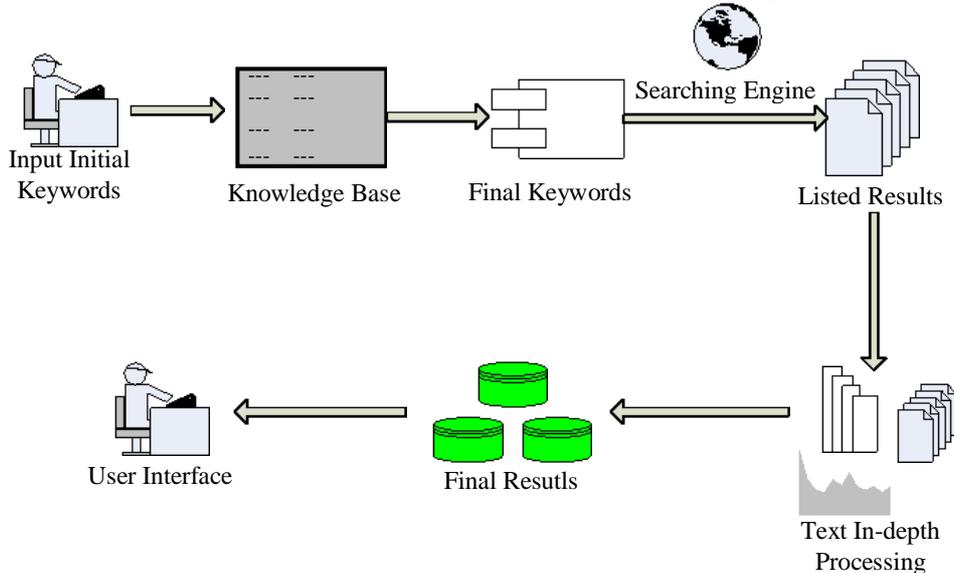


FIGURE 1 Framework of ISC

The transformation of keywords is completed through knowledgebase, in which keywords related to papers of similar categories are stored with table. A fragment is given as follows:

TABLE 1 A Fragment of Knowledgebase

Keywords	Category	Related Ones
term_1	C_1	term_1,, term_i
term_2	C_2	term_1,, term_j
.....
term_p	C_q	term_1,, term_m

Through this kind of transformation, users could expand or delete the keywords, and then strengthen the description about what they really want. For example, If we input "Stock" in any searching engine, thousands of results will come to us. To find out which one is more useful, we have to spend more time and energy. However, in knowledge base, "Stock" is related to "Corporation, Corporation Governance" in the field of "Research"; meanwhile, it is packed with "Trade, Commission, Bonus" together in package of "Finance". Users could fix on the category they are interested in, and with the help of knowledge base, "Stock" and other relative final keywords are input to searching engine, which will bring more specific results comparing to ones above. Some further text processing will be done to the listed results, including text classification, result format standardization, and so on. These are little associated to the technique we discuss in this paper, so no more description about them will be involved here. After the

processing of all steps in ISC, final results that are more concise and precise will be offered to users in the end.

3. ARTIFICIAL NEURL NETWORKS (ANN) IN ISC

3.1 About ANN

Artificial neural network is composed of a number of units, connected by weights, and it is well known about its performance on application, such as prediction, pattern recognition, machine learning, and classification. To build a neural network to perform a task, one must first decide how many units are included, how to initialize the weights of network, and which learning algorithm to be applied to a set of training and testing examples. The use of examples also implies that one must decide how to encode them in terms of inputs and outputs of the network.

3.2 Structure of ANN in ISC

As we have mentioned before, one of the crucial parts in ISC is how to build up a proper relationship of keywords in knowledge base. ANN is chosen as the optimization tool, and our purpose is to classify keywords according to category. Naturally, the input/output mapping in ANN is designed as keywords/categories, which determines the number of nodes in input and output layers with the layers being fully connected.

It's noticed that papers belonging to same category usually have different keywords to describe their attributes, which could be illustrated by Figure 2. Where

sample paper A has its unique keywords “U1...Um”, and together with sample B, they both have keywords series “S1...Sp”, meanwhile all the samples share the conjunct parts “K1...Kn”.

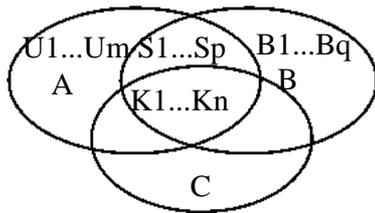


FIGURE 2 Keywords Vs. Papers

Apparently , these keywords have incremental importance in classification. Correspondingly, when a given category is set as the output, these weights connected to keyword inputs should have different value, that is the main purpose of network training.

3.3 Representation

The output data represents the class to which each paper belongs, that is expressed as 1 or 0 indicating the class to which a paper belongs or not. In view of input data format, it is a multi volume vector where 1 or 0 means certain training sample paper has or has not the correlative feature attribute. Take example we explained in the above section, suppose three papers belonging to same category have 15 keywords in sum, the input accordingly is a 15 dimension vector. For sample paper A, on the input volumes of keywords “U1...Um”, the value is set as 1, however about “B1...Bq”, the value is 0.

3.4 Learning Algorithm

According to the characteristics of this classification task, we choose MW-OBS, a pruning method, as optimization tool, which was put forward by us in a former paper. The basic idea of this method is that it could delete weights from a trained network, whose elimination induce so little increase in error that could be ignored. In this way, it could prune those unnecessary weights even nodes, if all weights connected to which are totally deleted. Reviewing the classification problem we described above, all keywords are set as the attribute input nodes. Through this kind of pruning, the keywords with dissimilar importance to some certain category would

have different, even none connected weighs to nodes of next layer, and that is why MW-OBS is chosen as learning algorithm.

3.5 Flowing Chart of MW-OBS in ISC Optimization

In order to illustrate how ANN is trained to satisfy the need of classification, a brief flowing chart is brought forth in next page (Figure 3) considering about the main purpose of optimization in ISC—to find out which keywords are crucial to the category. Based on such information, a knowledge base about keywords could be obtained.

Where E1 is a boundary value, preventing any elimination from inducing great increase in error that can't be ignored, in other words, no more weights could be deleted. Meanwhile M denotes the number of input nodes, and I as well as J is index variable. N(I) is a parameter that represents the number of weights connecting between Ith input node and nodes of next layer. E2 is another boundary value for distinguishing crucial ones from all keywords. When error increase L(IJ) is larger than E2, the Ith input is linked by important weights, whose elimination effects network error too much, and the Ith input is regarded as a pivotal attribute to output category.

4. AN ILLUSTRATIVE INSTANCE

Hundreds even thousands of materials would be included as training samples in order to build up a practical and comprehensive keywords knowledge database in extent or depth, which is a huge and systemic task, and is not to be described in details in this paper. Being part of our work until now, we introduce an illustrative instance listed in table 2 considering its use in ISC.

TABLE 2 An Illustrative Instance

Keywords	Category	Related ones
Basketball	Shopping	Basketball shoes
	Sports	Basketball match, Street, NBA
	Education	Basketball English, history
	Entertainment	Basketball games

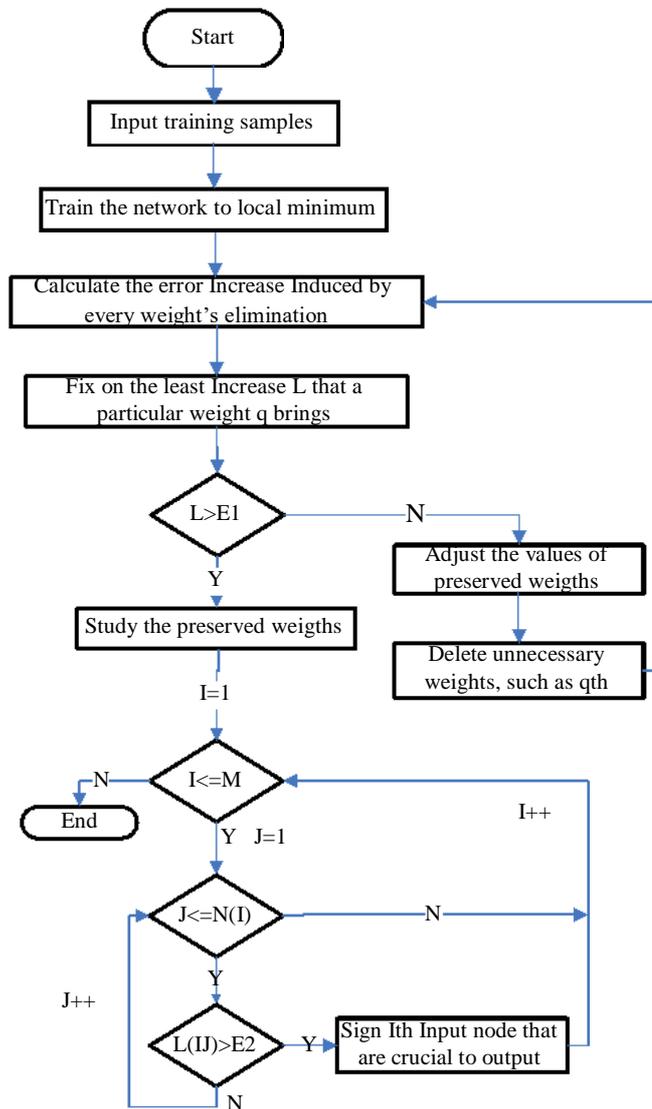


FIGURE 3 Flowing Chart of MW-OBS in ISC

About 200 paper samples are trained before arriving at these results. Reviewing the definition of “precision” offered in the previous section, several experimenters are invited to use this fragment of database to assist their searching process and give final judgments on results. Some statistical work shows that the average precision increases near 20% through the same searching engine since more accurate keywords are chosen for searching input, and most users are satisfied with the output of ISC.

5. CONCLUSION

The conception of ISC is put forward not only for the achievements in technical innovation, but also to emphasize the importance of coordination in information processing. Through integrating the separated parts into Information Supply Chain, higher efficiency and precision are obtained for users’ search.

With the application of knowledge base, users get final keywords, more specific and accurate descriptions than

original ones, which increase the searching precision, also help users to extend their comprehension about the related fields. Meanwhile, ISC, as a system, provides other utilities to users, such as defining logical relationship between keywords, text similarity calculation and classification. All these factors are organized and optimized in the purpose of satisfying the needs of users.

MW-OBS, a network pruning algorithm, unlike other optimization methods, deletes or adjusts the network’s weight according to their contributions to net error. In the topology of ANN after training, whether the weights connected to a certain node exist or not, as well as whether the values of them are large or little, directly decides this input node is the crucial keyword or not considering the classification task.

ISC, as a new information processing solution, has shown its advantages in primary experiment, and more related research and development of it will be carried out

in the future.

6. REFERENCE

- [1] LI Q, WANG Y. "An In-depth Study and Application Analysis of a Fast Network Pruning Approach: MW-OBS", *ICSSSM04*, Vol.2, pp. 620-623, 2004
- [2] Thomas J. Westbrook, J.D. "Integrating the Supply Chain", *www.wwwood.net*, 2002
- [3] Tyndall, G. R. "The Global Supply Chain Challenge". *Supply Chain Management Review*, Vol. 3, No. 4, pp.13-15.2000
- [4] Rich N., "Supply Chain Management and Time-based Competition: The Role of the Supplier Association", *International Journal of Physical Distribution and Logistics Management*, Vol. 27, No. 3, pp.210-225 1997
- [5] Wan W, Hirasawa K, Hu J, et al. "A new method to prune the neural network", *IEEE-INNS-ENNS Inter Joint Confer on Neural Networks*. Como, ITALY, 2000.
- [6] Boozarjomehry R.B.Svrcek W.Y., "Automatic design of neural network structures", *Computers and Chemical Engineering* 25, pp.1075-1088, 2001.

Study on Distribution Patterns of E-Business Based Agricultural Products Logistics

Yingrong Lu, Yinsheng Yang, Feng Lu, Xu Sun

School of Biological and Agricultural Engineering, Jilin University, Changchun 130025, China
yingronglu@163.com

ABSTRACT

The circulation of agricultural products is more complex and uncertainty than that of industrial products. The distribution channels of agricultural products are quite unique. The agricultural products logistics directly affect the efficiency of china's economy since agriculture is still main part of the economy. At the same time, with the development of information technology, e-business has affected people's daily life and brought great challenge to the traditional logistics of agricultural products. It is urgent to study the distribution pattern of agricultural products logistics under the circumstance of e-business. In fact, the platform of e-business provides a new approach for rationally deploying agricultural products. The agricultural products logistics system can be optimized based on the e-business-based distribution model. In this paper, we will focus on the logistics patterns and models of e-business-based agricultural products. After discussing the characteristics of the e-business-based logistics system, the model of logistics distribution based on electronic data interchange (EDI) and dynamic allocation mechanism will be presented. Considering the specific processes of agricultural products logistics, we depicted a framework called e-business logistics management (EBLM) model, which can integrate e-business functions and distribution functions. Finally, we will discuss the implementation strategy of EBLM.

Keywords: *agricultural products logistics, E-business, distribution patterns, EBLM model, implementation strategy*

1. INTRODUCTION

The characteristic of the information society is the optimization of information flow and IT-based innovation. The development of the Internet provides such a tool that is changing the market rules and forming the new network economy. As the representative of network economy, E-business has already been applied in nearly every aspect of our social activities, such as retail, service, manufacturing, etc. Especially, Internet has changed the operation behaviors of enterprises thoroughly. Annam, the general secretary of the United Nations, said that e-business is pushing the economy development, improving the trade efficiency, and enhancing the competition ability of enterprises. It can not only integrate the national economy into the world economy, but also can provide the employment opportunities and create more wealth.

Although we still cannot predict when the e-business dominated trade will become the primary trading mode, the potential of the e-business market is endless. According to the U.N. E-business Development Report in 2002, the number of global Internet customers has reached 6.5 billion by the end of 2002^[1]. In China, the number has reached 87 million by the end of June in 2004 according to the latest investigation report from the China Internet Network Information Center (CNNIC). On the other hand, the volume of e-business trade grows fast. Forrester, a famous consultation company, estimated that the volume of global e-business trade in 2002 is about 2,293.5 billion USD, and it will attain 12.8 trillion USD in 2006, which will be 18% of the world retail volume, and the average growth rate is

above 30% each year. In China, The U.N. E-business Development Report in 2001 indicated that the volume of B2B and B2C market would be 4 billion USD in 2003, B2B would increase 194% each year, and B2C will be 274%. Due to intensive labor force in agriculture, the development of e-business of agricultural industry is more difficult than other industries. Although the agricultural circulating network has developed at certain level in our country, and some regions such as southeast area has already started to construct the e-business-based agricultural products system, there are no systematic planning because of lacking corresponding theory. A lot of research works have been done about seasonal products^{[3][4]}, but there is still no special concern about agricultural problems. For the information-based agriculture, it is very important to improve the circulation of agricultural products under the e-business environment. That is why we will study the e-business-based agricultural products logistics patterns in this paper.

2. THE FUNCTION OF THE LOGISTICS DISTRIBUTION IN E-BUSINESS DEVELOPMENT

The logistics distribution is a very important link that completes the whole e-business-based trade.

2.1 Modern logistics is the basis of e-business

Each e-business trading process generally needs three essential elements: the logistics, information flow, and cash flow. The logistics is the foundation; the information flow is the bridge; and the cash flow is the

purpose. Each e-business trading process usually accompanies with the logistics and information flow. The trading partners need information so as to send out, follow, divide, receive, store, pick up and pack the goods, etc. In information era, the cooperation of logistics and information flow becomes extremely important, and must be supported by modern logistics techniques.

Logistics techniques refer to the general name of all professional techniques related to logistics elementary activities, including various working techniques and management skills such as circulating machining technologies, packing technologies, identification technologies, real time tracing technologies etc. With the development of computer network technologies, logistics technologies synthesized a lot of modern information technologies, for instance, GIS, GPS, EDI, Bar code, etc. The logistics industry accelerates to use the modern information technology. This will pave the way for the popularization of the electronic commerce.

2.2 The logistics distribution system is the supporting system of the e-business

Modern logistics distribution can provide service for the customer of the e-business. According to the characteristics of e-business, it can implement the unified management of the information, schedule the whole logistics distribution system, complete to manage the goods at the request of customers, and send goods to the consignee. This logistics mode is significant to logistics enterprises for improving the service quality, reducing the logistics cost, raising enterprise's economic benefits and social benefits.

2.3 The logistics distribution system improves the efficiency of the social economy circulating

The logistics enterprises adopt computer network technology, application software systems and advanced management theory, according to the customer's demand, they carry on a series of arrangements, such as classifying, allocating, ordering, dividing, mixing goods, etc., to deliver all kinds of goods in right time, right place and right quantity. The new logistics model becomes the pioneer in the field of circulation, and represents mainstream of modern marketing. On the other hand, it is easier to realize informationization, automation and socialization in the new logistics system than in the traditional one. It can not only reduce the stock of manufacturing enterprises, accelerate the turnover of the fund, improve logistics efficiency, reduce the logistics cost, but also can stimulate the social demand, promote the sound development of the economy.

3. ANALYZING THE CHARACTERISTICS OF AGRICULTURAL PRODUCTS LOGISTICS DISTRIBUTION

3.1 The characteristics of agricultural products logistics distribution

After China re-entered world trade organization (WTO), the bulk agricultural products of China, such as cereals, oil, cotton and sugar, will be in a inferior position in international market and the production cost is obviously higher than the same agricultural products price in international market. But for the labor-intensive agricultural products, such as vegetable, meat, birds, eggs, as well as aquatic products, it will be in a superior position both in amount and price. The market mechanism of agricultural products has been a certain scale in our country. We focus on the study about logistics system of agricultural products here.

For agricultural products, the logistics chain is depicted in Figure 1.

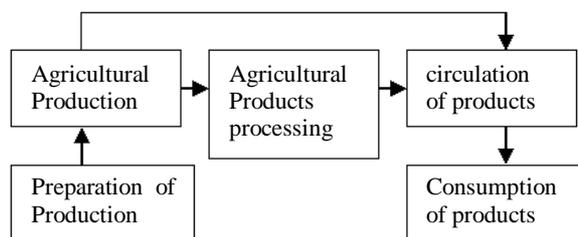


Figure1 agricultural products logistics chain

In Figure 1, the preparation of production includes planting seed industry, feed industry, information guidance, products planning, etc.; agricultural production includes field management, instructor, fertilizer planning, etc.; agricultural products process includes grade classify of products, packing, fresh-keeping, storage, etc.; the circulation of products includes channel planning, a network of commercial establishments laying out, circulation technology of cold chain, market building of product and sale, etc.; consumption of products includes providing enough variety and quality of goods, guiding consumption, etc..

Agricultural products logistics chain includes all links and procedures from cultivating seed to field management, agricultural products processing, keeping fresh until circulation, market sell. Each link on the chain is related with many other industries, and such basic elements as agricultural science and technology, agricultural information and standardization etc. are influencing the agricultural industry chain, the distribution network that takes agricultural production as the main thread will be formed finally. What it pursued is to form various mechanisms to realize the goal of resource rational distribution.

The characteristics, such as localization, seasonal cycle, scattering etc., are very outstanding, meanwhile, the agricultural products are people's everyday necessities, the consumption elasticity is low.. The several characteristics exist in the whole logistics include:

1) The dispersion characteristic of produce distribution link

The producing area of agricultural products is very wide. The production unit, such as farmers, does not understand the market information and lack of essential production planning, thus the production link becomes the most difficult to control in the whole logistics chain. Dispersion characteristic of production leads to dispersion of the front logistics in the industry chain.

2) The region characteristic of circulation distribution link

Agricultural products are widely produced in different regions, and people's demand is diversified, therefore circulate trade occurs in different areas. But because agricultural products are fresh and perishable, even if we take measures to keep fresh, they will still have some proportion loss, and this proportion will be strengthened with time and distance running up. With the circulate cost rising, the radius of the agricultural products circulation is restricted. On the link of the fresh products delivering, keeping cold and keeping fresh of the fresh products and the cycle of the process have restricted in the supporting service radius of the center. This is obviously different from the normal temperature logistics.

3) The specified distribution facilities

Because the agricultural products are fresh and perishable, we must take some measures in circulating, so that agricultural products could be consumed with high quality. Some variety specific container and equipment will be needed^[2] in the course of agricultural products warehousing and transportation. The cost of agricultural products logistics distribution should be higher than the traditional industrial products, so extensive and specialized management will be needed.

4) Integrative characteristic of delivering and trade (the trading process accomplished by distribution)

In recent years, our country builds 3-level marketplaces for agricultural products circulation, namely the wholesale market of agricultural products in producing region, the wholesale market in selling region and the retail market of farm and produce. The 3-level wholesale markets take not only the function of the trade but also the functions of the delivery. It is showed that agricultural products logistics will be difficult to exist independently when it break away from the trade center.

We can see from these characteristics that the agricultural products logistics chain is different from industry's logistics relying mainly on production, its key link lies in circulation that will drives the operation of the whole logistics chain .

3.2 the key problems and solutions of agricultural products logistics distribution

Our country has always pays attention to agricultural products circulation system. According to the characteristics of the agricultural products logistics distribution, our country comes on quite loose policies and regulations for encouragement to set up modern agricultural products logistics system. As the foundation of agricultural products logistics is relatively weak, we still face a lot of problems now. Considered the whole logistics chain, the key problems among them include:

1) Strategy positioning of the agriculture products logistics

In the field of agricultural products circulation, there is still lack of the concept of the whole logistics system at present. The operator on each link is often engaged in the commercial activity taking oneself as the core. In terms of the logistics strategy of the whole industry, the end of logistics chain-consumption- is emphasized. Therefore the development of agricultural commerce centers and fresh food supermarket are fast, but it can't drive other links in the chain.

2) Lacking flow of the information in the logistics chain

Rational flow of the information is even more important in the efficient logistics chain, that is, there should be a valid information share mechanism among each link in Fig. 1, which can improve the operation efficiency of each link. But so far as the present situation in our country, mechanisms which can collect, store, analyze and use information of agricultural products logistics chain have not been set up. The incompleteness and dissymmetry of information results in blindly and inefficient production. Poor information also leads to the existence of a large amount of invalid logistics. According to statistics, agricultural products' losses in the course of circulation are probably 75 billion to 80 billion every year in China, accounts for 20% of the output value.

3) Lacking the unified planning of logistics distribution system

Region and dispersion characteristics of agricultural products make it is difficult to manage and plan logistics distribution. Up to now, we have not set up rational planning method yet. Trans-regional fresh products logistics, because its loss is heavy and its costs are high, whose scheduling is more unable to realize.

For these questions, the government of China has responded correctly, six government ministries jointly point out in 2003: the agricultural products circulation facilities are an important component of the social infrastructure. And the report further defined the strategic position of agricultural products circulation. Under the guidance of this policy, we can solve three key problems mentioned above from three following aspects:

1) Focusing on circulation, build up strategy of the

agriculture products logistics chain

The idea of agricultural products supply chain management should be spread. At the same time, the traditional logistics chain, which taking consuming as the core, should be transformed. Actually, circulation link is the most potential part of the supply chain, so the new strategy should be based on circulation, and the function of the circulate link should be enhanced correspondingly. In detail, based on the traditional collecting and distributing center and new e-business environment, the circulation part should has functions such as production guiding, trading and information management etc. The strategy with circulation as the core will fully utilize e-business environment, contribute to overcome the disadvantage of decentralized agricultural production.

2) Improve information collection and analysis among logistics chains, strengthen the information function of the traditional collecting and distributing center (CDC)

The communication and coordination of information are important foundation of the modern logistics management. Information management in the CDC should be further strengthened after the center of the logistics chain is transferred to the circulation link. The e-business trade platform of agricultural products will be set up on the basis of setting up two levels collecting and distributing system, namely, CDCs in producing area and selling area. On one hand, the information of production, demand and relevant information on the logistics chain will be classified, stored and shared. On the other hand, the trans-region scheduling of the

agricultural products will be realized by the information analysis function. Based on the analysis of demand information, the agricultural production will no longer be blind also.

3) Based on e-business environment, set up new management systems of so-called e-business-based integrated agricultural products logistics
E-business technology and the derived methods of logistics management is the foundation of our new logistics strategy. On the basis of fully considering the characteristics of logistics and information flow in the course of agricultural products circulation, we will set up an integrated e-business-based agricultural products logistics management models.

4. E-BUSINESS-BASED LOGISTICS MANAGEMENT (EBLM) MODEL FOR AGRICULTURAL PRODUCTS

As we mentioned above, the key of solving agricultural products logistics problems is to establish an e-business-based platform that logistics and information flow are considered and integrated. This platform has not only traditional wholesale functions, but also new Internet-based functions such as electronic trade, logistics management, information share, balance of supply and demand and production planning, etc. Aiming at this goal, we will establish an E-Business-based Logistics Management (EBLM) Model for Agricultural Products (figure 2).

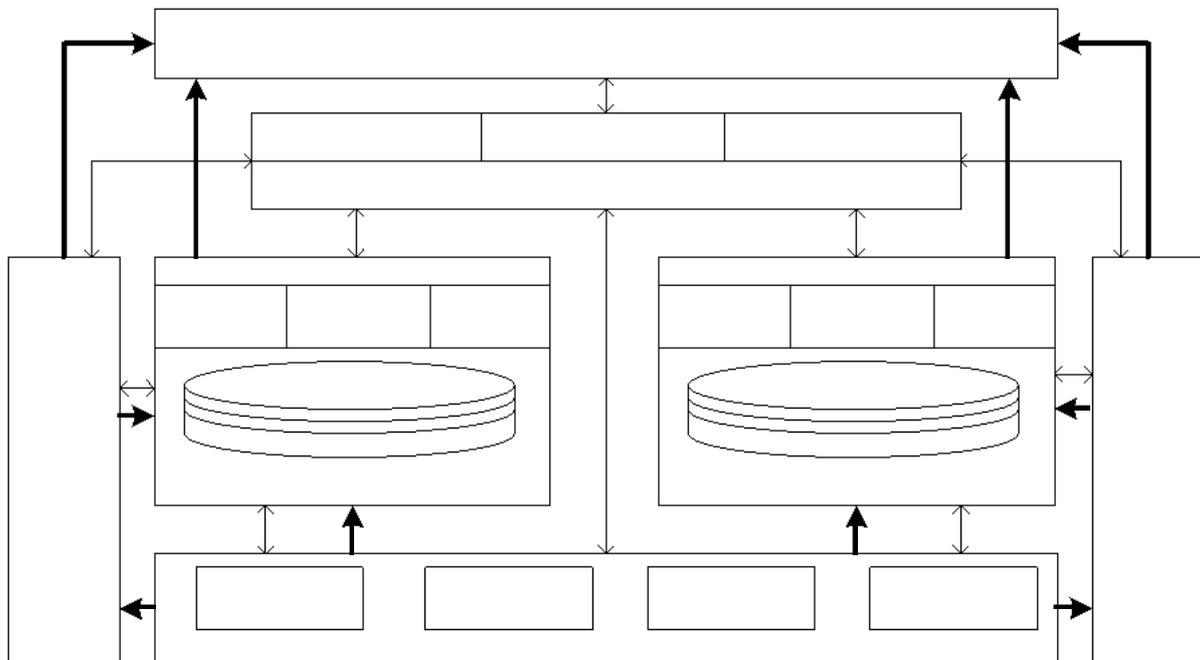


Figure 2 E-business-based Logistics Management Model for Agricultural products

In figure 2, the wide arrows indicate logistics and narrow arrows show information flow. The EBLM model is

composed of 3-level supply chains: agricultural products processing unit, distribution & wholesale center and retailers. The third party logistics will be fully used to fulfill transportation of agricultural products among 3-level supply chain. The functions such as information share, online trade, and schedule optimizing and market prediction will be realized by setting a platform of e-business.

Specifically, agricultural products processing unit is made of farmers, processing factories, farms and importers, which are headstream of logistics and production information provider. As noted previously, agricultural products processing unit in our country has been dispersed, but if we set up the portal system of the regional distribution center, information can be exchanged conveniently on the portal among these processing units. This will solve various kinds of problems because of scatterance. In addition, processing unit will emphasize pre-treatment of the agricultural products according to the demand information, which will effectively lower invalid logistics.

Distribution & wholesale center (DWC) is physical center and information node. It combines the functions of traditional wholesale markets both in producing area and sale area. Different centers may have different functions according to the characteristics of the physical region of the center. Meanwhile, in this center, the two databases: agricultural products database and logistics database, and a portal website will be set up also. In the portal website, the third party logistics can be scheduled and agricultural markets trends will be analyzed. In addition, the center is the hub of downward information flows, and agricultural products processing units can get direction from the analyzing function of the portal.

The retailer is composed of farm produce marketplaces and supermarkets. It is the end of EBLM and faces to consumers directly. The retailers get agricultural products through the coordination of e-business platforms, distribution centers and the third party logistics. An important function of retailer in the model is the collection of the market information. On the basis of the statistical analysis of sales information, it fixes the general market trend and formulates the corresponding purchase plan, then offers information to the e-business trade platform.

The e-business platform is a virtual center of EBLM, which can be operated independently or based on a certain DWC. The platform connects all DWCs to realize information share in each link of agricultural products supply chain, schedule trans-region logistics, deploy limited transportation ability, such as rational utilization of cold chain circulates ability etc..

5. IMPLEMENTATION STRATEGY OF EBLM

The EBLM model has compressed the length of the

traditional agricultural products supply chain by the virtual and physical centers, and strengthened the functions of information share and analysis among every links of the supply chain. This model can improve the efficiency of the agricultural products circulation effectively, reduce circulation cost and contribute to realize strategic idea to set up of new agricultural products circulation system. However, the foundation of the application of agricultural information technology is still very weak, the third party logistics is not developed enough in China either, so it's difficult to implement the EBLM.

In order to implement EBLM smoothly, firstly, we should do the unified planning of the agricultural products circulation facilities, and should prevent repeated construction. The government in policy and in certain regular range should guarantee the unified standards about implementation of the EBLM. The unified planning will contribute to preventing from malignant competition and setting up normal environment of agricultural products circulation.

Secondly, the construction of the agricultural products market's information network should be strengthened. The Ministry of Agriculture has already done the work in this aspect, some agricultural products information networks are running and releasing information now. But the functions of these systems are limited and need further improvement. On the request of EBLM, the agricultural products e-business platform is relatively centralized and simple, so we should adopt the standard component, thus it can lower costs by a large and popularize rapidly. Another focus is how to set up database system, including the mechanism of data collecting, processing, storing and using.

Thirdly, the rational implementation tactics step by step should be constituted. The infrastructure of agricultural products logistics should quickly be constructed at present to promote trans-region circulation of agricultural products and cross-seasons supply and to guarantee the market supplies. Agricultural products logistics in main producing region and main demanding region should quickly be set up, especially for western food to eastern area. The peasant's cooperative transportation organization and large management enterprises should be fostered to guide actively and encourage peasants to transport and sale agricultural products together. This is the basis of forming DWCs. The agricultural products management enterprises should be supported to build some circulation facilities to preserve transport agricultural products. We should strengthen the ability in transport and selling, expand business scale and develop made-to-order farming to improve the organization efficiency in China. Based on the analysis above, the cooperative society and the agricultural products management enterprises should be turned into progressively distribution center of EBLM, these are DWC.

Fourthly, the multi-channels social funds for the development of EBLM are encouraged. Government will encourage, enterprises and individuals to attend agricultural products circulation system construction. The trans-region development logistics distribution and catenation management should be encouraged. The current management system of wholesale marketplace should be reformed to set up the new operating mechanism in which the real investors will get the real benefits, and they will assume sole responsibility for its profits or losses.

Finally, the circulation should open to the world. The involvement of foreign partners can improve the transportation capacity and management level of the logistics industry in China so as to promote the implementation of EBLM model.

In addition, the state's policies should be fully utilized. Our country will arrange some allocating funds every year during the Tenth Five-Year Plan period to subsidize the information system of agricultural products market, pesticides residues monitoring system and electronic settlement and infrastructure such as the roads, marketplaces, etc.. The local government must also offer certain related funds to the project that the central government supports. In the view of this point, The National Development & Reforms Committee issued a document called "the notice on printing and distributing pilot project implementation suggestion of wholesale market of agricultural products of 2003"^[1] in April of 2003, In this document, central government will arrange more than 400 million RMB national debt funds to support 81 projects. Majority of them is the marketplace, others are agricultural products storing and fresh-keeping enterprises. At the same time, the government will continue implementing the supporting policies, such as bank loans, finance pays interest etc.. Taking the year of 2003 as an example, projects on agricultural products circulation, including HuaLian and LianHua supermarket in Shanghai, gained more than 10 billion RMB loan with low interest.

Based on EBLM, agricultural products circulation system can be set up by the country's support, enterprise's participating and rational implementation tactics.

6. CONCLUSION

With the development of information technologies, there

is a strong tie between modern logistics and e-business. It is important to develop an integrated e-business-based logistics system. In this paper, the problems of the e-business-based logistics of agricultural products are explored, and the characteristics of agricultural products logistics distribution are analyzed also. In order to solve the problems in the traditional agricultural products logistics, the e-business-based agricultural products logistics management model (EBLM) is put forward and the implementation tactics are discussed also.

Agricultural products logistics in China has great potential. The operation efficiency of the whole agricultural products supply chain will be improved effectively by setting up e-business logistics system with circulation as the core. Under the guidance of this strategy, implementing EBLM model can carry on effective management to logistics and information flow in the agricultural products supply chain so as to set up normal and efficient agricultural products circulation network.

In order to implement EBLM properly, a systematic approach should be adopted, and get the support from the governments, enterprises, and farmers.

Further research will focus on the analysis of the issues involved to the EBLM from macro and micro views, for example, the study of macro policies of EBLM, the construction scheme of e-business platform, information standard and share mechanism, etc..

ACKNOWLEDGEMENT

This paper is partly supported by the Science and Technology Ministry of Jilin Province and Jilin University. Grant number is 20040528.

REFERENCES

- [1] Geng shuhai, Recent policies for agricultural products circulation, www.cafte.gov.cn
- [2] Fang Xin, Status and trends analysis of fresh supply chain, www.emkt.com.cn
- [3] Chen, J., L. J. Xu, "Coordination of the Supply Chain of Seasonal Products", *IEEE Transaction on Systems, Man and Cybenetics: Part A*, 31(6), 2001, pp524-532.
- [4] Chopra S., P. Meindl, *Supply Chain Management: Strategy, Planning, and Operation*, Prentice Hall, Upper Saddle River, New Jersey, 2001.

Study on Supply Chain Management in Electronic Commerce

Yuhua Hou¹, Zheng Sun²

¹College of Management, Hebei University, Baoding 071002, China

²Humanity College, Agriculture University of Hebei, Baoding 071000, China
hyhua60825@163.com

ABSTRACT

Electronic commerce has been changing the operation models of firms to adjust the ever more competitive environment in the the commercial field, and the new ones of cooperation amonge firms have been formed by it. In the globalized electronic commerce, the old and traditional models of supply chain management have fade away by being failed to fill the needs in the electronic commerce, and the new models of, therefore, should be created. Firms strive to adapt their supply chains to the electronic commerce, and they need to decide how much and in what sequence they will be investing in various supply chain efforts to leverage the benefits of the new technology. In this paper three main aspects are described. First, based on the present a comprehensive definition of supply chain management, relevant characters of it that have risen to the importance with the prevalence of electronic commerce are provided. By using the universal criteria in the Internet, the information in many directions being flowed has been already realized, so, the commercial partners can build a automatic and integral supply chain, and the chain can operate as a whole. Second, the strategies of implementing suply chain management are discussed. By applying the supply chain management in electronic commerce, the time of communication between companies has been shorten; the nicety has been realized; and the efficiency of stock and conveyance has been greatly improved. Moreover, by integrating multi-enterprise, the implementations of concurrent engineering and inventory-reinforce plan have been allowed.

Key Word: Supply Chain Management, Characters, Strategies, Electronic Commerce

1. INTRODUCTION

The rapid growth and adoption of the the electronic commerce have already had a great impact on all aspects of business, including customer acquisition, marketing, human resource management, finance, information systems (IS), and operations. Supply chain management that has played an important role in traditional businesses is likely to be crucial in the electronic commerce environment as well^[1]. To survive in the competition, the companies have, especially for someone who compete in the low technical-obstacles industries, have to be closed connected with whole-sellers, dealers and retailers to reduce the cost of products and to gain to competitive edge. At during the next five years, collaboration by supply chain partners over the electronic commerce can potentially save \$223 billion with the reduction in transaction, production, and inventory costs^[2]. This could be one of the largest benefits of electronic commerce technology and many firms have already begun to realize these benefits. For example, Microsoft used a Web collaboration tool to bring the Xbox video game console to market two months ahead of schedule^[2]; Cutler-Hammer has doubled profits and increased productivity by 35% for its configured motor control centers and control panels by coordinating with customers over the electronic business^[3]; and Autoliv reduced the plant inventory by 37% by coordinating orders online with suppliers^[4]. The large potential impact of the electronic commerce on supply chain management makes its study timely and important.

2. SUMMARIZE OF SUPPLY CHAIN MANAGEMENT AND THE E-COMMERCE

2.1 The Conception of Supply Chain Management

Having a concrete understanding of supply chain management, which has been enriched in the trend of globalization of production and service, is the precondition to success of operating it. Then, what is the supply chain? A supply chain is the set of entities (such as whole-sellers, producers, retailers, transportors, and final consumers etc.) involved in the design of new products and services, procuring raw materials, transforming them into semifinished and finished products, and delivering them to the end customer. Supply chain management is the efficient management of the end-to-end process, which starts with the design of the product or service and ends with the time when it has been sold, consumed, and finally, discarded by the consumer^[5]. This complete process includes product design, procurement, planning and forecasting, production, distribution, fulfillment, after-sales support, and end-of-life disposal. The utility of supply chain management is to boost the cooperation between many entities, including whole-sellers, producers, retailers, transportors, and final consumers. The cooperation, which enables to optimize all the relations concerned with the supply chain, thus, provides the biggest value to buyers by minimizing the cost.

It is clear that supply chain management spans several functional and geographical areas, introducing

complexities both in terms of design and execution. Some of the pertinent factors that complicate supply chain management decisions include the presence of multiple agents and their sometimes conflicting incentives; uncertainty in demand, supply and production distribution process; asymmetry of information related to product design, inventory, costs, demand and capacity across the supply chain between the various parties in the supply chain; and lead time between the different entities in the supply chain. All these complexities lead to several types of inefficiencies in the supply chain: poor utilization of inventory assets^[6]; distortion of information such as the bullwhip effect^[6], high stock-outs and nonresponsive supply chains^[7]; poor customer service due to customization-responsiveness challenges^[8]; and double marginalization that leads to lower profits for the supply chain^[9]. Setting up models of supply chain management for electronic commerce in the ever developing environment is the guarantee of successful implementation.

2.2 The Conception of Electronic Commerce

Why electronic commerce can be loosely defined as IC (Internet Commerce) is because the electronic commerce process is developed by the technologies, mainly by using EDI and the Internet or other electronic medium as a channel to complete business transactions^[5]. Therefore, all the ways of exchanging information, which was not used to be transformed by Internet, are completely changed by doing it in the supply chain. They enables the scarce resources to be fully applied, the commercial taches and time period to be abridged, so, the cost cutting, efficiency rising, quality reinforcing are all realized.

Electronic Commerce consists of three areas^[6]: (1) consumer-oriented activity and (2) business-oriented activity supported by (3) the electronic commerce technology infrastructure. The technology infrastructure relates to network infrastructure, network applications, decision technologies, and software tools and applications.

3. CHARACTERS OF SUPPLY CHAIN MANAGEMENT IN ELECTRONIC COMMERCE

Internet has created opportunities to integrate information and decision making across different functional units, thereby creating a new channel for companies and buyers in supply chain that go beyond a business unit to study the extended enterprise^[5]. Supply chain management has coordinated the relationships among departments and the members in the supply chain. By applying more simple, fast, and much lower cost E-communicating way contacting whole-sellers, retailors, producers, and buyers, the relationships among departments and members in the supply chain

have been coordinated. Thus, electronic commerce has greatly elevated the role of supply chain management and analysis within a firm, and made it with the unique characters.

3.1 The Information Sharing Fully

Internet has also opened up opportunities for firms to share information and efficiently coordinated their activities with other entities reinforced in the supply chain. Integration of important data in the supply chain by fully sharing information is the guarantee to provide undistorted and timely information.

3.2 The Synchronization of Supply Chain

In the supply chain management process, companies establish more effective ways of E-communication by applying E-mail, E-meeting, E-marketing, and database integration, and intagrating information systems among firms to make sure that all the nodes in the supply chain are accessed information at the same time. Therefore, the synchronization of the supply chain is realized.

3.3 The Integration of Supply Chain Management

Supply chain, with producers, whole-sellers, retailors and buyers, enables companies and other related organizations connected into a net, in which all units form a concrete strategic relationship based on sharing the universal benefits and harms^[10]. Since the shared information is unlimited by time, space and the increasing flow of it, the pattern of information transaction has become a net too. Thus, all the units in the net can trust eath other more, and form a strategic alliance. The integration of suply chain management, therefore, is boosted.

3.4 The Innovation of Business Process

Supply chain managemen in electronic commerce has changed many aspects of business process from purchasing to logistics^[11]. This change needs every units in the supply chain, especially the core members, to operate the internal BPR (business process reengineering) from a macro-coordination. To realize the changes of business process, innovations have been made to adjust the needs of higher efficiency and flexibility for supply chain managemen.

3.5 The Speed of Operating Suply Chain Management

Being back up by the Internet technologies, the supply chain management enables all members in the supply chain to reduce the unnecessary and nonprofit steps of the process to cut the cost. This is the ensuring to the realization of promptness and fineness.

4. THE STRATEGIES OF IMPLEMENTING SUPPLY CHAIN MANAGEMENT IN ELECTRONIC COMMERCE

To make sure the success of implementing the supply chain management, a suit of strategies should be established.

4.1 The Establishing Basic Components for Supply Chain Management

Supply chain management is the combination of art and science that goes into improving the way in which your company finds the raw components^[12]. Supply chain management needs to make a product or service, manufactures that product or service and delivers it to customers^[13]. The following are five basic components which are established for supply chain management.

4.1.1 Plan-This is the strategic portion of supply chain management. You need a strategy for managing all the resources that go toward meeting customer demand for your product or service. A big piece of planning is developing a set of metrics to monitor the supply chain so that it is efficient, costs less and delivers high quality and value to customers.

4.1.2 Source-Choose the suppliers that will deliver the goods and services you need to create your product or service. Develop a set of pricing, delivery and payment processes with suppliers and create metrics for monitoring and improving the relationships. And put together processes for managing the inventory of goods and services you receive from suppliers, including receiving shipments, verifying them, transferring them to your manufacturing facilities and authorizing supplier payments.

4.1.3 Make-This is the manufacturing step. Schedule the activities necessary for production, testing, packaging and preparation for delivery. As the most metric-intensive portion of the supply chain, measure quality levels, production output and worker productivity.

4.1.4 Deliver-This is the part that many insiders refer to as "logistics." Coordinate the receipt of orders from customers, develop a network of warehouses, pick carriers to get products to customers and set up an invoicing system to receive payments.

4.1.5 Return-The problem part of the supply chain. Create a network for receiving defective and excess products back from customers and supporting customers who have problems with delivered products.

4.2 The Process of Implementing Supply Chain Management in Electronic Commerce

4.2.1 Setting Up the Companies' ERP/MIS System. The units in the supply chain, especially the core members, having the mature informational system, means that all the members in the supply chain have the smooth of informational flow and the internal business process, and that they can apply the connecting system of the upstream facility and the downstream facility to ensure the realization of smooth the business process. The mature and concrete informational system in a company, therefore, is the precondition for the successful implementation of supply chain management in electronic commerce.

4.2.2 Making Coordination Project between Upstream Facility and Downstream Facility in the Supply Chain. Once the project is begun, its team should be organized by all units to confirm the technological criterion, platform, exchanging necessary data and database format together. For different units have different database format and different products' code, a transforming system according to them should be set up to let all units to be freely accessed to the utility of the supply chain. After the coordination project and the timetable are built, every unit in the supply chain is required to develop their Web Service according to itself criterion.

4.2.3 Applying the BPR for Logistics or Business Process to Meet the Requirement of Supply Chain Management in Electronic Commerce. How much the supply chain management could benefit a company depends on how much the operating models of the company could adjust the informational needs. The desired effects of the supply chain management in electronic commerce could not be embodied, even the benefit would be cut off by the failure of coordination between both sides, if the systems between upstream facility and downstream facility in the supply chain are connected only, but the business process with it is not changed. For this reason, a full analysis of business process should be done firstly, the unefficient process deleted and integrated, and the business process redesigned to meet the need of coordination of the supply chain.

4.2.3.1 The Innovation of Business Process within A Firm. Helped by the advanced information technologies and the thoughts of supply chain management, the internal business process of a firm has made great change^[11]. Take the relation of the business process between producing departments and purchasing department as examples. Under manual manipulation conditions, the model of sequacious engineering was applied, in which the production plans made by the firm's managers, programed by the departments to supply raw materials, asked for permission from many layers of an organization, and sent the order form to suppliers after above works done. Because the prolonging process and involving too many departments and people form out of joint, holdup and repeat, a

project would always be done through a long time. In the supply chain management, however, the original business process can be synchronized its steps and largely simplified by sharing database in a high-tech platform of information, which makes the original model of sequacious engineering transformed to concurrent engineering. For instance, the purchasing supply department can make its decision promptly by acquiring the plans according the information on the database after producing departments finishing the producing plan, by counting the raw materials cosumed necessarily and suitable parts. Furthermore, the order form can be sent directly to the suppliers through Internet or EDI in the supply chain.

Anyway, the innovation of internal business process is to meet the needs of the coordination in the supply chain, and to realize the informatinal management of whole process in the line, which embodies its effects by the intensivism of information, by the simplifying and the adjusting a firm's internal business process, and by reducing the middle-level management and many complex taches.

4.2.3.2 The Innovation of Business Process among Firms. Under the circumstance of applying supply chain management, the business process has been made much easier by the application of Internet or EDI, for many business taches done by manure labor in the past are made unnecessary or even erased with such the more modern way as E-commerce. For example, suppliers were used not to prepare for producing until receiving the order form from buyers, and after the parts being ready, so much time had been wasted. The whole producing process would be much prolonged in such a vicious circle. Under supply chain management, however, all firms are accessible to the freshest information of the production schedule to make better and timely decisions for production, supply, and delivering with Internet. At the same time, a supplier who knows the condition of consumption on the buyers' side by being informed from Internet, is capable to shorten the period of supply goods, especially during the time that the stocks are full, , and makes a full preparation before not receiving the orders from manufacturers so that the cycle is largely shortened. Because of this more effective way of coordination, the original departments, positions, and processes, which are for the jobs that can be simplified, can be considered to re-design.

The innovation of business process among firms based on the innovation of business process within a firm alter ulteriorly the business process among members in supple chain so that responses to the customers are shortened and the stronger feeling of satisfaction will be realized.

4.2.4 The Strategy of Integration Web Service Concerned with Firms to the System of Connection.

The connection between the upstream facility and downstream facility in the supply chain based on the Internet makes the concerned firms to freely exchange information, and the works of debugging and test followed the exploitation of the system operated.

4.2.5 Acquiring the Support and Cooperation from the Parners. Different from the innovation of business process within a single firm in supply chain, the innovation of business process among firms are much more difficult for changing not only the business process within a firm but also other firms', including operating interosculation and keeping touch with suppliers and distributors. Therefore, in desired condition, the innovation of business process of supply chain management should be set out from a macro prospective, and all the nodes should be synchro-reformed. In other word, the whole harmony among units, especially jointing firms in the supply chain, is the most important point, though the situations in different firms are different. Otherwise , the coordination of the supply chain would not be promised.

4.2.6 The Evaluation to the Supply Chain Management in Electronic Commerce. It is necessary to harmonize the operation models among all units of the supply chain , and to adjust the data platform to meet the realistic need in time by the evaluation to the supply chain management.

5. CONCLUSION

With the rapid development of electronic commerce and the fierceness of competition, supply chain management will have brought ever greater profits on firms^[14], provided better services and more greater values for customers, and griped more worthy opportunities of the new market in order to make give the firms an advantage over in the fierce competition^[15]. The paper shows the characters of the supply chain management in electronic commerce, which can fulfill the requirements of low-cost and flexible, and the models are applicable for many companies. As long as the problems of safety, financial ability and technology could be ensured settlement, supply chain management in electronic commerce will be made sure to embody the imponderable advantages, especially in sharing information and concurrent engineering, to be comparable to the traditional management of supply chain, and become a prevalence in the near future as a new and exciting management model^[16].

REFERENCES

- [1] Geoffrion, A., G. W. Graves, "Multicommodity distribution system design using benders decomposition", *Management Sc*, Vol. 20, NO. 5, pp 822-826, 2001.
- [2] Keenan, F., S. Ante, "The new teamwork", *Bus. Week*, (February 18), 2002.

- [3] Bylinski, G, "The e-factory catches on", *Fortune*, (July 23), 2001.
- [4] Lundegaard, K. Bumpy ride:, "Supply chain management sounds beautiful in theory; in real life, its a daunting task", *Wall Street J*, (May 21), 2001.
- [5] Swaminathan, J.M.,S.R.Tayur, "Models for supply chain in E-Binessin", *Management Sci.* Vol. 49, NO. 10, pp1387-1406, 2004.
- [6] Lee, C. Billington, "Managing supply chain inventories: Pitfalls and opportunities", *Sloan Management Re.* Vol. 33, NO 3, pp 65-77, 1992.
- [7] Fisher, M. L., J. H. Hammond, W. R. Obermeyer, A. Raman, "Making supply meet demand in an uncertain world", *Harvard Bus. Rev*, Vol. 72, NO. 3, pp 83-92, 1994.
- [8] Swaminathan, S. R. Tayur, "Managing broader product lines through delayed differentiation using vanilla boxes", *Management Sci.* Vol. 44, NO. 6, pp161-172, 1998.
- [9] Cachon, G. "Supply chain coordination with contracts", S. Graves, T. de Kok, eds. *Supply Chain Management—Handbook in OR/MS*, Vol. 11. North-Holland, Amsterdam, The Netherlands, 2003.
- [10] Wang Liujie, "Integration of supply chain", *WWW.NET.CN*
- [11] Ma Shihua, The organization and the innovation of business process in supply chain management", <http://www.56zg.com/books/scm/11-3.htm>
- [12] Huggins,E.L.,Olsen,T.L,"Supply chain management with guaranteed delivery", *Management Sci*, Vol. 49, NO. 9, pp1387-1406, 2003.
- [13] Kulp,S. C., Lee, H. L.,Ofek,E, "Mannufacturer benefits from information integration with retail customers", *Management Sci.* Vol. 50, NO. 4, pp 431-444. 2004.
- [14]Li Huayan, Ma Shihua, Lin Yong, "The coordination andperformances of supply chain", *The World of Electronic*, pp100-106, 2003, 2-
- [15]Lin Yong, Ma Shihua, "A theoretic study on cooperation and principal applied in supply chain corporation", *Nankai Business Review*, pp49-53, 2000, 2.
- [16] Sun Jian, Lin Zefu, "The Impact of electronic commerce to China's M&S enterprises", *Nankai Business Review*, pp42-44, 2000, 2.

Supply and Demand Management under Inducement of Price Discounts – A Monte Carlo Simulation Analysis

Henry C. Co

College of Business, California State Polytechnic University, Pomona, CA 91768, U.S.A.
Tel: 1-909-8692732, Fax: 1-909-8692466, Email: hco@csupomona.edu

ABSTRACT

This paper considered a single-item, three-echelon (supplier, retailer, and customer) inventory problem. At random times, the supplier offers the retailer a discount. The inter-arrival times of discounts are exponentially distributed. For the retailer, whether or not to take a discount offer depends on its inventory level. If inventory is below threshold level S , the retailer will order to replenish inventory up to $S+Q$ units. Otherwise, the retailer will pass. At the regular purchase price, the retailer uses the order-up-to reorder point inventory system. The reorder point is r , and the order-up-to quantity is R . The demand of the customer is price-elastic. Depending on its inventory level, the retailer may discounts its selling price to boost up demand. The decision variables are S , Q , R , r , and the selling prices (regular and discount selling price). The simulation experiment covered reasonable range of values for each decision variable. Each run represented one scenario, i.e., one combination of the decision variables.

Keywords: demand management, price discount, multi-echelon inventory system, simulation

1. INTRODUCTION

Consider a three echelon (supplier, retailer, and customer) inventory system. At random times, the supplier offers the retailer a price discount. The inter-arrival time of discounts is exponentially distributed with expected value of 100 days.

For the retailer, whether or not to take the discount offer depends on its inventory level. If inventory is below threshold S , the retailer will order to replenish inventory up to $S+Q$ units. Otherwise, the retailer will pass.

At the regular purchase price, the retailer uses the order-up-to reorder point inventory system. The reorder point is r , and the order-up-to quantity is R . The demand of the customer is price-elastic. Depending on its inventory level, the retailer discounts its selling price to boost up demand.

In this paper, we used Monte Carlo simulation to investigate the following scenarios: (1) the basic model, (2) the limited-time offer model, (3) stochastic extension of the limited-time offer model, (4) pricing to clear inventory, and (5) stochastic extension of pricing to clear inventory.

The decision variables are S , Q , R , r , and the selling prices (regular and discount selling price). The simulation experiment covered reasonable range of values for these decision variables. Each run represented one scenario, i.e., one combination of the decision variables.

In the simulation analysis, the ranges of values for the decision variables were:

- $1,000 \leq S+Q \leq 7,000$
- $50 \leq R \leq 3000$

- $50 \leq S \leq 500$
- $\$7.00 \leq \text{regular selling price} \leq 10.00$
- $\$5.00 \leq \text{discount selling price} \leq 8.00$

The cost parameters used in the simulation were:

- Unit Cost (Regular) $C_I = \$5.00/\text{unit}$
- Unit Cost (Discount) $C_D = \$4.00/\text{unit}$
- Holding Cost/Year 15% of cost
- Ordering Cost \$50
- Backorder Cost \$10 /unit

Each combination of the decision variables was simulated for 365 days, and replicated as many times as necessary for making meaningful statistical inferences. Descriptive statistics on the contribution margin for each scenario was generated. The contribution margin is the difference between annual revenue and annual relevant cost. The annual revenue equals daily sales in units * selling price for the day, for 365 days. The annual relevant costs include item cost, ordering cost, holding cost, and shortage cost. The item cost is based on the weighted average unit cost.

Clearly, the weighted average unit cost is lower if the retailer orders more often at the discount price.

The results of the simulation were plotted. On each plot, a polynomial trend line was fitted based on the points on the graph.

All computations were made using Microsoft Excel on a Pentium 4 2.6 GHz computer.

2. THE BASIC MODEL

The assumptions for the basic model are:

- Demand rate d is deterministic, i.e. $d = \alpha p^{-\epsilon}$, where $\alpha=1,000$ and $\epsilon=3$.

- At random times, the supplier offers the retailer a discount. The inter-arrival times of discount offers are exponentially distributed with mean = 100 days. **The discount offer is good until the retailer takes the offer.**
- The retailer offers discounts to its customers whenever the inventory level is above the threshold level S . Below S the selling price reverts to the regular price. Since there are two selling prices, there are two levels of demand rate (demand is price elastic) Each run consisted of 1,000 replications. Each 1,000 replication took less than 3 seconds. The results of the simulation are summarized in Charts 1-1 to 1-5 below:

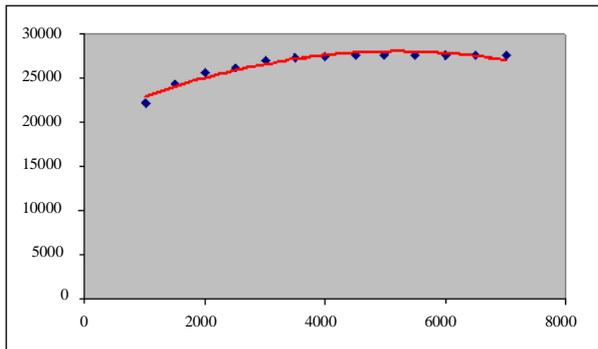


Chart 1-1 Contribution Margin vs. $S+Q$

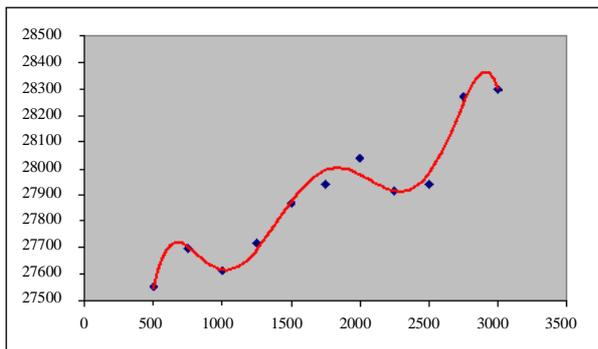


Chart 1-2 Contribution Margin vs. R

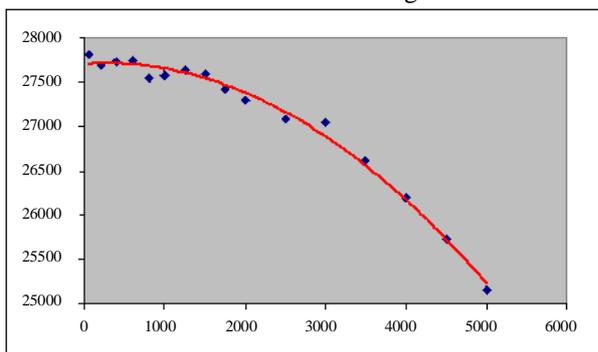


Chart 1-3 Contribution Margin vs. S

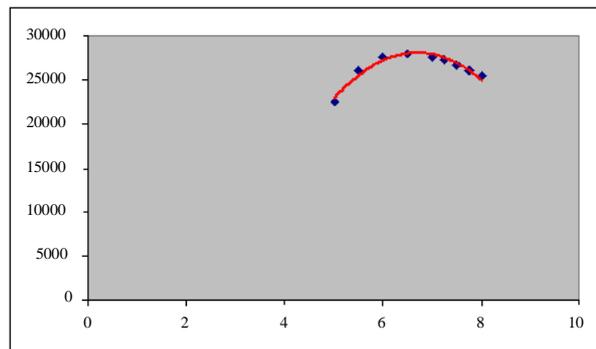


Chart 1-4 Contribution Margin vs. Discount Price

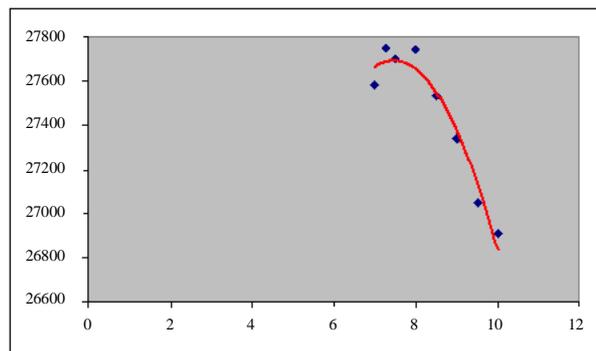


Chart 1-5 Contribution Margin vs. Regular Price

The contribution margin appears to be rather flat in the vicinity of $S+Q=6,000$. The trend-line for the Contribution Margin vs. R chart (Chart 1-2) appears to be roller-coaster-like. However, from $R=500$ to $R=3,000$, the difference in contribution is less than \$750. The trend-line in Chart 1-3 shows the contribution margin dropping very fast beyond $S=1,000$. The trend-lines in Charts 1-4 and 1-5 suggest a regular selling price of \$7.50 and a discount selling price of \$6.50.

3. LIMITED-TIME-OFFER MODEL

In the basic model, each discount offer from the supplier is valid until the retailer takes the offer. This is not common in practice. In the ensuing models, we considered the more realistic practice of "limited-time-offer." The Limited-time-offer Model retained all other assumptions in the Basic Model, except that each offer of discount is good for 5 days only.

The following charts show the results of the simulation for the Limited-time-offer model:

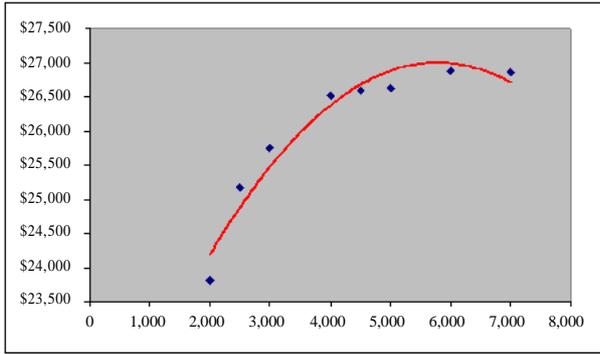


Chart 2-1 Contribution Margin vs. S+Q

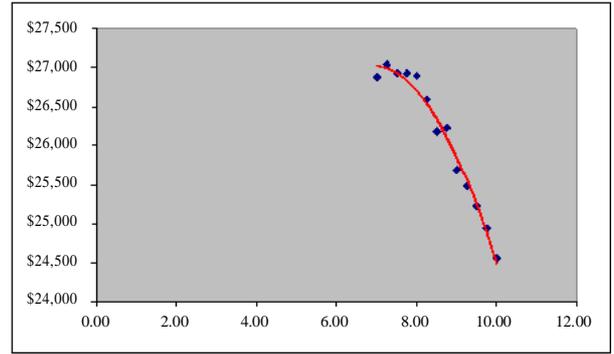


Chart 2-5 Contribution Margin vs. Regular Price

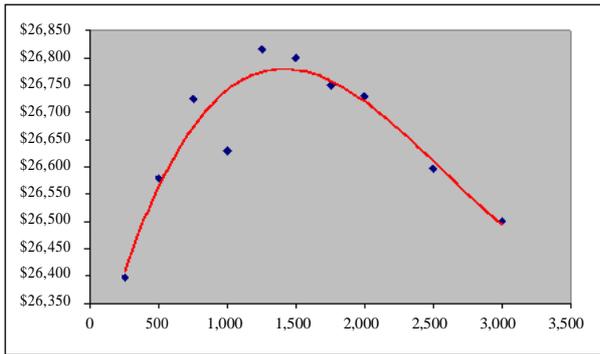


Chart 2-2 Contribution Margin vs. R

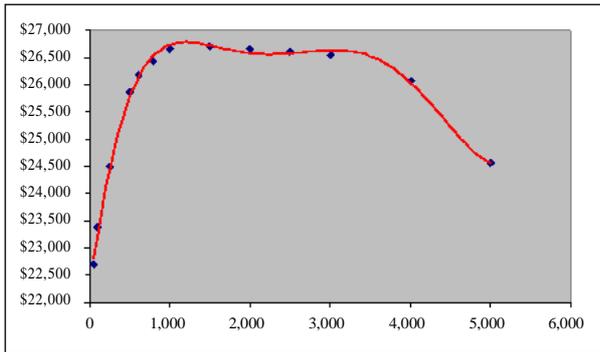


Chart 2-3 Contribution Margin vs. S

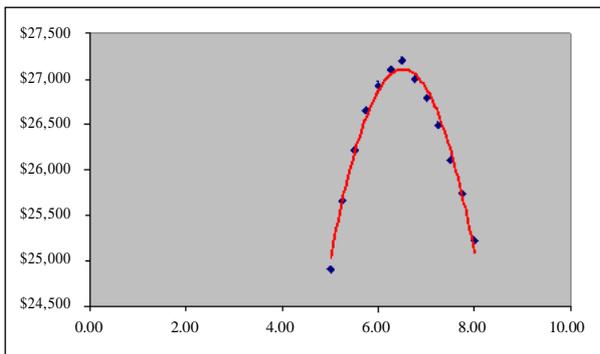


Chart 2-4 Contribution Margin vs. Discount Price

The trend-lines appear to be unimodal. The table below shows the summary statistics, based on the input parameters suggested by the charts above:

- Threshold Level S=3,000
- Order Quantity R=1,500
- Max Inventory S+Q=6,000
- Nominal Price \$7.50
- Discount Price \$6.50

<i>Contribution Margin</i>	
Mean	27,583
Standard Error	63
Median	28,151
Mode	20,817
Standard Deviation	2,002
Sample Variance	4,009,841
Kurtosis	0.545
Skewness	-0.858
Range	9,375
Minimum	20,817
Maximum	30,192
Sum	27,583,363
Count	1,000
Confidence Level (95.0%)	124

4. STOCHASTIC EXTENSIONS

In this model, we extended the above scenario and considered stochastic demand. During the lead-time, inventory level may drop down to zero before replenishment arrives (a shortage situation), or the level may be positive when replenishment arrives. Thus the reorder point r is another decision variable of interest. We examined two alternative probability distributions – (a) Truncated Normal (i.e., demand is non-negative) and (b) Exponential.

The trend-line charts for this scenario are very similar to the ones shown above, and will not be reproduced here.

5. DISCOUNT TO CLEAR EXCESSIVE INVENTORY.

Since demand is price-elastic, the retailer can clear excessive inventory by adjusting the selling price accordingly. In this model, we analyzed a scenario where the retailer tweaks its selling price periodically, based on inventory level. This model is very common among sellers of commodity items or perishable goods. For example, a bumper crop of grape leads to overproduction of wine, bringing down the price of wine in retail stores. Overproduction of oil and petroleum drives down prices. As reserves are depleted, the prices slowly rise upwards. This is also very common among grocers.

This model retained all other assumptions in Limited-time-offer model except that retailer adjusts the selling price based on beginning inventory. The % discount is selling price is $I^{(1-\delta)}/100$, where I = beginning inventory, and $\delta = 0.6250$.

The following charts show the results of the simulation for the Limited-time-offer model:

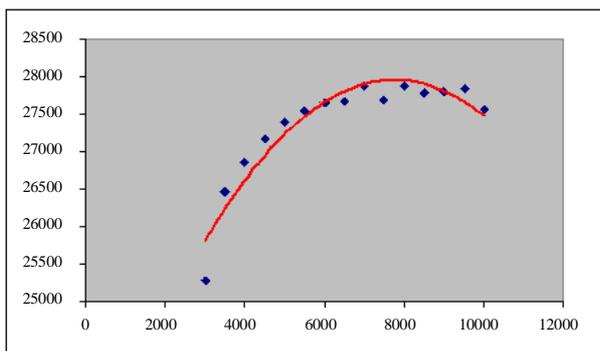


Chart 3-1 Contribution Margin vs. S+Q

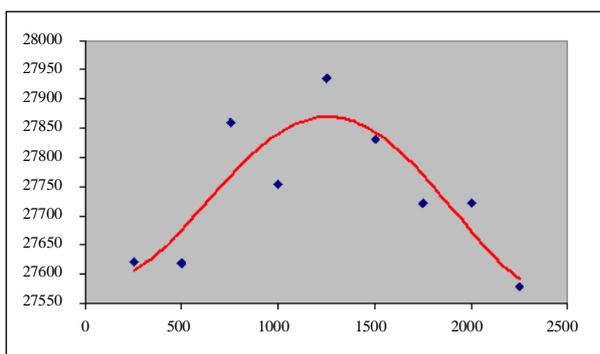


Chart 3-2 Contribution Margin vs. R

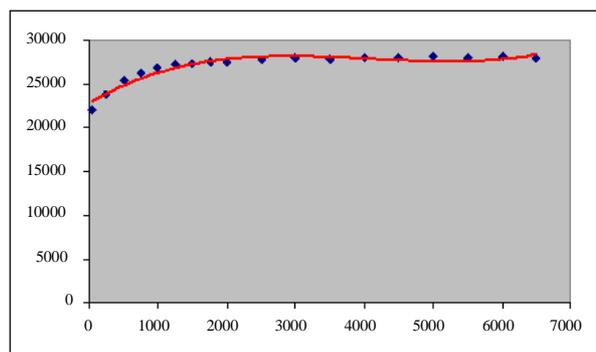


Chart 3-3 Contribution Margin vs. S

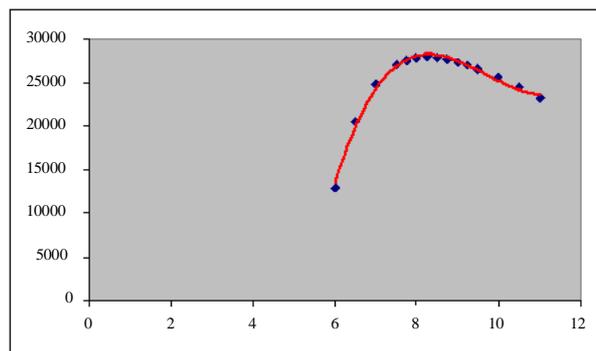


Chart 3-4 Contribution Margin vs. Regular Price

The trend-lines appear to be unimodal. The table below shows the summary statistics, based on the input parameters suggested by the charts above:

- Threshold Level S 3,000
- Reorder Point r 60
- Order Quantity R 1,500
- Max Inventory' (S+Q)7,000
- Nominal Selling Price\$8.00

<i>Contribution Margin</i>	
Mean	27,839
Standard Error	74
Median	28,094
Mode	20,389
Standard Deviation	2,352
Sample Variance	5,530,168
Kurtosis	0.835
Skewness	-0.898
Range	10,673
Minimum	20,389
Maximum	31,062
Sum	27,838,767
Count	1,000
Confidence Level (95.0%)	146

6. DISCOUNT TO CLEAR EXCESSIVE INVENTORY: STOCHASTIC EXTENSIONS

In this model, we extended the preceding scenario for exponentially-distributed demand. The expected demand $d = \alpha p - \epsilon$, where $\alpha=1,000$ and $\epsilon=3$.

The trend-line charts for this scenario are very similar to the ones shown above, and will not be reproduced here.

7. CONCLUDING REMARKS

This paper demonstrated the application of Monte Carlo simulation to analyze the problem of supply management under inducement of price discounts. We investigated the three echelon (supplier, retailer, and customer) inventory problem under the following scenarios: (1) the basic model, (2) the limited-time offer model, (3) stochastic extension of the limited-time offer model, (4) pricing to clear inventory, and (5) stochastic extension of pricing to clear inventory.

The decision variables are S , Q , R , r , and the selling prices (regular and discount selling price). The simulation experiment covered reasonable range of values for each decision variable. Each run represented one scenario, i.e., one combination of the decision variables. The results of the simulation were plotted. On each plot, a polynomial trend line was fitted based on the points on the graph.

Examining the trend-lines, we were able to identify appropriate values for the decision variables. The author is currently working on an analytical approach to determine the optimal values of these decision variables.

REFERENCES

- [1] Abad, P. L. (1997), "Optimal policy for a resupplier when the supplier offers a temporary reduction in price," *Decision Sciences*, 28, No. 3, 637-653.
- [2] Baker, R.C. and T.L. Urban (1988), "A deterministic inventory system with an inventory-level-dependent demand rate", *Journal of the Operational Research Society*, 39, 823-831.
- [3] Friend, J. K. (1960), "Stock control with random opportunities for replenishment," *Operations Research Quarterly*, 11, 130-136.
- [4] Goh, M. and M. Sharafali, (1998), "Price-dependent inventory model with discount offers at random times", *Research Paper Series RPS 98-43*, Faculty of Business Administration, National University of Singapore.
- [5] Golabi, K. (1985), "Optimal inventory policies when ordering prices are random", *Operations Research*, 33, 575-588.
- [6] Hurter, A. P. and F. C. Kaminsky (1968a), "Inventory control with random and regular replenishment," *Journal of Industrial Engineering*, 19, 380-385.
- [7] Hurter, A. P. and F. C. Kaminsky (1968b), "Inventory control with a randomly available discount purchase price," *Operations Research Quarterly*, 19, 433-444.
- [8] Moinzadeh, K. (1996), "Replenishment and stock policies for inventory systems with random deal offerings," *Management Science*, 43, No. 3, 334-342.
- [9] Taylor, S. and C. Bradley (1985), "Optimal ordering strategies for announced price increases," *Operations Research*, 33, 312-325.
- [10] Tersine, R. J. and E. T. Grasso (1978), "Forward buying in response to announced price increases," *Journal of Purchasing and Materials Management*, 20-22..

Taking Advantage of E-Logistics to Strengthen the Competitive Advantage of Enterprises in China

Jianjun Wang, Deli Yang, Qiong Guo, Yunfu Huo

School of Management, Dalian University of Technology, Dalian, Liaoning 116024, China
wmeagle717@sina.com

ABSTRACT

With the fast development of the global economy, how to deliver the right products to the end-users in the demanding time and places becomes more and more important to the enterprises. In order to improve the competitive advantages, Chinese enterprises, especially the large manufacturers, need to develop e-logistics function and modern supply chain management systems. This paper describes the changing environment, such as competition, global trends, and logistics operations. Then the e-logistics is discussed, and it presents that the Descartes can improve the level of the information sharing and creates a new source of advantage to ensure competitiveness in today's dynamic manufacturing marketplace. At last some suggestions are given, including establishing an integrated network, selecting the right industries, and so on.

Keywords: E-logistics, supply chain, information system

1. INTRODUCTION

Now China has become the manufacturing powerhouse in the world. With the development of the global economy, the vast majority of goods are bound for global destinations. China produces the majority of the world's televisions, VCD, refrigerators and other goods. China's share of world trade is already in the top 10 nations and has an annual economic growth rate of almost 8% the past decade. In 2003, China accounted more for 3% of the global trades; some experts forecast it will be 10% in the 2020 (Shawn P. Daly, Lindsay X. Cui.2003).

Fueling Chinese industry is an increasing number of multi-billion dollar Chinese enterprises with manufacturing, distribution and services on the whole country or global scale. Many of these enterprises have regional manufacturing capabilities and distribution, warehousing and service facilities close to customers in the whole country, even in Japan, North America, Europe and other parts of the world.

2. IMPACT ON COMPETITION

Since the advent of internet technology, the computing environment in many applications has been changed drastically. The increased information requirements have facilitated an integration of logistics information systems (LIS) and supply chain information systems in many companies. The increasing use of electronic commerce and enterprise resource planning and other logistics information systems tools and techniques will shape the business process for the foreseeable future. Companies should understand their options and their impacts when making decisions to support their supply chain systems (Stephen M. Rutner.2003).

Of the four key elements to succeed in business -- product, price, promotion and place -- manufacturers nowadays are able to produce the same quality of products at the same price using the same promotions. However, they have developed a different strategy for place -- the channel of distribution to deliver goods and services from factories to end-users. But China's logistics infrastructure is straining under the weight of its new economic growth. Chinese companies have begun to realize that E-logistics is the new frontier of competition in the digital economy.

3. GLOBAL TRENDS AND LOCAL IMPLICATIONS

Most large-scale Chinese companies have developed strategic partnerships with global information technology leaders such as SAP, IBM, HP, Dell, Gateway, Ericsson, Motorola, NEC, and many others. As a result, Chinese companies are becoming increasingly integrated into global supply chains that must be agile, flexible and -- with the adoption of the internet -- electronically linked to ensure they are responsive to the needs of today's rapidly changing and demanding marketplace.

Accelerating this integration is the rapid adoption of various Internet technologies by global technology leaders as they work to sustain and strengthen their competitiveness in today's digital marketplace. These initiatives tend to fall into two broad categories (1) supply chain management; (2) customer relationship management

3.1 Supply Chain Management

To send products that customer needs to the retailer and the final customer quickly becomes the main means of the enterprise's surviving and competitive advantage.

The competition not only happens between the firms, but also happens in the different supply chain in the future. The main task of supply chain is supporting logistics from the primitive supplier, several processes of produce and transport to customer finally (Niklas Aldin.2003).

Companies large and small are launching a broad range of supply chain management initiatives designed to improve operating efficiencies, raise service levels and boost profitability. Initiatives such as B2B integration, e-procurement, collaborative planning and – more recently – e-logistics are being launched in cooperation with trading partners to increase efficiencies and drive waste out of supply chains.

3.2 Customer Relationship Management

Customer Relationship Management (CRM) can manage the customer's information effectively and meet their needs of production and service, in order to attract new customer, lock the frequent customer for enterprises and improve benefits and competitive advantage it develops long-term, steady and trusting relationship with the customer.

More and more firms communicate with their customer by information technology under the environment of e-business, which makes it possible to extensively collect and analyze the customer's information. It resolves the problem of customer information management, which puzzled the enterprises for a long time, it also improve the customer's service level and customer's satisfaction. It obtains and keeps the customer for the enterprise; i.e expends and keeps the financial resource for the enterprise. So the customer relationship management has already become the important content of the organization management innovation under the e-business.

Every communication between the firm and the customer is regarded to be beneficial and value-added in the CRM, the customer's loyalty to the firm is won by mutual communication with the customer. Through correctly guiding, most customers would like to share their own information with enterprises (Nianlian Yi.2003 [6]).

CRM mainly involves such "enterprise's front desks" as market intelligence analysis, marketing and customer service etc, in order to complete it requires the positive cooperation of customer-oriented functional departments. Meanwhile, if it wanted to offer satisfactory products and service to the customer, the processes such as product design, manufacture and produce controlling etc which are regarded to be the "enterprise's backstage supporter" should be involved. The idea and practice of CRM in the e-business need the cooperation of the function department of the whole

enterprise; it also needs the adjustment and innovation to the whole enterprise's procedure.

Companies are adopting e-business to develop deep, personalized relationships with their customers (often consumers). Nearly 50% of Dell's \$32 billion in annual revenues are now transacted over the web and other manufacturers and distributors are also reporting sharp increases in on-line sales. And the U.S. is not the only country with booming e-business sales. In Japan, for example, e-business is expected to grow from \$7 billion in 2001 to \$25 billion by 2003, with approximately 40% of this from PCs and peripherals alone.

4. IMPACT ON LOGISTICS OPERATIONS

Both types of initiatives – supply chain management and customer relationship management – dramatically impact how China suppliers and their foreign business partners manage their global logistics operations. The initiatives highlight the need for real-time, accurate logistics information that can be shared by supply chain players to allow for more intelligent logistics decision-making and timely responses to planned and unplanned events in the supply chain.

E-business in particular is accelerating the way logistics are managed by companies across the entire value chain. The size of shipments are shrinking, shipment frequencies are increasing and the ubiquity of the internet is providing new opportunities – and challenges -- for companies serving customers who are geographically dispersed, difficult to predict and sensitive to price and service levels.

Table1 compares traditional logistics with e-logistics of today (Yaohua Song, Hanping Hou.2003).

Table1. The Difference between Traditional Logistics and E-Logistics

	Traditional Logistics	E-Logistics
Shipment Type	Bulk	Parcel
Customer	Strategic	Unknown
Customer Service	Reactive, Rigid	Responsive, Flexible
Distribution Model	Supply-driven Push	Demand-driven Pull
Inventory/ Order Flow	Unidirectional	Bidirectional
Destinations	Concentrated	Highly Dispersed
Demand	Stable Consistent	Highly Seasonal, Fragmented
Orders	Predictable	Variable

E-business has a ripple effect up the supply chain that impacts China assemblers and their suppliers as foreign buyers work to synchronize their operations with actual customer demand. Historically, supply chains have been structured to "push" products from suppliers to consumers with limited – if any – information available

about real-time consumer demand. This “push” approach typically results in supply chain distortions marked by excess inventories, obsolete products, unsatisfied consumer demand and an inability to forecast accurately.

On the other hand, “pulling” products across the supply chain helps to ensure supply and demand are synchronized by feeding information about actual - not forecasted - orders to the enterprise, suppliers and third party logistics providers. The benefits of a pull system are obvious but realizing them requires fundamentally rethinking how logistics should be managed throughout the entire order to delivery process.

Now, Chinese companies and their customers gradually begin to rely on third party logistics providers (such as freight forwarders) to coordinate the logistics and related decision-making required to ship products from main land factories to various destinations. However, due to industry fragmentation and technology limitations it has been difficult (if not impossible) to obtain accurate information on the position and status of goods as they move across the supply chain. As a result, the logistics operations of most shippers, customers and logistics providers themselves function in an extremely information poor environment despite the fact that logistics plays a critical role in their business operations: the timely and cost-effective delivery of goods to customers.

5. E-LOGISTICS

E-Logistics is a dynamic set of communication, computing, and collaborative technologies that transform key logistical processes to be customer centric, by sharing data, knowledge and information with the supply chain partners. E-logistics also enables synchronization of events and right decision-making. The ultimate objective is to deliver the right products in right quantities at the right place and time to the right customers.

E-Logistics leverages the power of the internet and other technologies (such as wireless) to provide robust information to supply chain participants and offer unprecedented levels of visibility across the entire supply chain. A growing number of e-logistics solution providers and service companies are tapping into this opportunity by addressing logistics issues such as supplier selection, asset utilization, pricing, inventory management, order visibility, and order fulfillment. Start-ups, software companies, and “old economy” logistics providers are bringing to market a variety of products and services that address the huge logistics inefficiencies burdening today’s supply chains.

While these companies may indeed provide value to Chinese shippers, the Descartes can provide the level of global connectivity required to create value through

networked communities. The key attractions of this networked community are (1) fewer integration points required to connect to network members; (2) an exponential increase in value for each member as more participants join the community; and (3) a common infrastructure for the communication and exchange of information. Imagine for a moment a world without a global telecommunications network where instead phone users strung a separate line to each new person they wished to speak with.

The Descartes’ Global Logistics Services Network (GLSN) provides technology to ensure consistency and integrity of communication throughout supply chain communities. It is an information platform encompassing thousands of transportation carriers, logistics intermediaries, and users of transportation services around the globe. The network securely registers, connects, and manages participants. In addition to providing a standard for inter-enterprise business connectivity, it acts as a logistics process framework for end-to-end shipping management.

The GLSN enables organizations to build high-performance supply chains, providing real-time visibility and decision support. Its open architecture allows trading partner systems to seamlessly integrate with logistics-critical data and business rules, wherever and however they reside. It enables organizations to connect with a single, simple mechanism that incorporates security, so only preauthorized users access information, applications and documents. The GLSN manages the integration of the hundreds of data communication standards, and maintains responsibility for data monitoring, quality and integrity. And recognizing that each participant may have different technological abilities and needs, it supports all common communication protocols, leveraging comprehensive embedded dictionaries for translation.

Tapping into this global network can provide Chinese shippers and their customers with visibility of orders (not just shipments) across their supply chains. Manufacturers can use the Descartes network to “link” together their manufacturing order systems with shipment tracking systems to provide a single, unified view into the status of orders as they make their way from PO to final delivery. With true order visibility, shippers, buyers, and logistics providers understand where goods and inventories reside across the supply chain and can make intelligent, proactive logistics decisions based on this information

Beyond the obvious benefits of order visibility across the supply chain, access to accurate, real-time logistics information provided by the Descartes Network allows companies to undertake various improvement initiatives that can drive savings, increase revenues and improve their overall competitive advantage. Based on experts’ estimates and experience, companies undertaking such

initiatives can realize a 10-20% reduction in their total logistics costs, improve productivity and enhance service levels.

Logistics improvement initiatives enabled by Descartes include inventory management, shipment planning, mode selection, reverse logistics optimization, performance measurement and other efforts. In the past, such efforts have been extremely burdensome to implement due to limited logistics data. However, armed with meaningful logistics information, management can bring their logistics operations out of the dark ages and into new era. This will improve the overall responsiveness of their supply chains and create a new source of advantage to ensure competitiveness in today's dynamic high-tech manufacturing marketplace.

6. SUGGESTIONS

6.1 Establishing an Integrated Network

Looking at it from a bigger picture, the current logistics landscape is fragmented. Therefore, the key challenge is for us to ally together to establish a proper network of services to cover the whole region efficiently. It is an arduous challenge to pick out the right partners to work with in providing an end-to-end supply chain solution. Further to that, our country alone has very different customs, cargo and tax policies. Therefore, it is vital to engage local partners in order to overcome any barriers involved in clearing overseas shipments.

6.2 Improving the Awareness Level Of Logistics Services

The low awareness level of high-value online logistics services is a most pressing issue to deal with at present in the market. There are a lot of enterprises in China that are not fully convinced of the value of adopting IT-enabled processes in their operations let alone e-logistics services. Creating the much-needed awareness is very important in order to keep the business flowing. Actually, customers need to view e-logistics services as the means not just to lower costs and shorten time cycles but, more importantly, to enhance their competitiveness in the marketplace by concentrating on their core competencies.

6.3 Selecting the Right Industries

Another addressed challenge is to pick the right industries to target. Each industry has its unique culture and requirements and thus it is difficult for a service firm to cater to a wide range of industries. The firms need to focus on building up expertise in particular market segments so as to capitalize on particular market opportunities.

6.4 Establishing Training Mechanism

With regards to available resources, traditionally the manpower in industry is of a physical nature or manual-labor-intensive. Incorporating IT into the services requires a higher level of skills. This type of skill set is lacking at the moment. Thus, the challenge is to establish training facilities to increase the availability of skilled and qualified labor to man the IT-enabled processes.

6.5 Building an E-Logistics Product Portfolio

Local and foreign clients have different requirements and cultural expectations. Although the current demand stems more from foreign clients, it is important not to overlook the potential demands of locals. Hence, a key challenge is to be able to create a product portfolio that can encompass the needs of both sets of clients. Furthermore, it is already hard enough to cope with the ever-changing needs of client requirements. Turning all the customer needs into an integrated and flexible set of services is a tall enough order even for the most experienced industry players. A good way to understand the customer needs thoroughly is to embrace the customers' business challenges as one's own.

6.6 Improving the Technology Level

It is inevitable that companies with strong financial backing or relevant industry experience will have an upper hand in developing sophisticated technology for their clients. As such, newly established companies will find it difficult to compete with them in building a solid set of services. In this industry, credibility and reputation are valued assets to have. Research has revealed that even a simple but important technology such as bar coding is still missing in most warehouses in the whole country. The technology gap between the local and international players is also a key concern. So we should narrow the technology gap between the local and international players as soon as possible.

6.7 Industry, Government and the Academic Working Together

In order to take advantages of e-logistics, the industry, government and the academic should work together.

The industry should focus their effort on the following issues:

- ò Adoption
- ò Technology development
- ò Investment and integration
- ò Adoption and development of processes.

The issues that belong to the government are as follows:

- ò Awareness development
- ò Standards setting
- ò Regulatory oversight
- ò Intergovernmental relationships
- ò Infrastructural support and development.

Academia should take the following aspects into their research fields:

- ò Basic research and development
- ò Implications of technology on the environment
- ò Determining what works and what doesn't
- ò Educating and publicizing into their research fields

7. CONCLUSION

Today's supply chains require effective logistics management to meet the requirements of increasingly demanding global customers able to choose when, how and where they want to receive their goods. The urgency for improvement is being accelerated by the widespread adoption of internet technologies and applications such as e-business in the global market. E-logistics can improve the overall responsiveness of supply chains and create a new source of advantage to ensure competitiveness in today's enterprises.

ACKNOWLEDGEMENT

This paper is supported by the National Natural Science Foundation of China (No: 70031020).

REFERENCES

- [1] Shawn P. Daly, Lindsay X. Cui, "E-logistics in China: basic problems, manageable concerns and intractable solutions", *Industrial Marketing Management*, Vol. 32, No. 3, pp235-242, 2003.
- [2] Stephen M. Rutner, Brian J. Gibson, Susan R. Williams, "The impacts of the integrated logistics systems on electronic commerce and enterprise resource planning systems", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 39, No. 2, pp83-93, 2003.
- [3] Niklas Aldin, Fredrik Stahre, "Electronic commerce, marketing channels and logistics platforms—a wholesaler perspective", *European Journal of Operational Research*, Vol. 144, No. 2, pp270-279, 2003.
- [4] Nianlian Yi, "Research of new management mode of e-business organization", *Commercial Research*, No. 5, pp129-131, 2003.
- [5] Yaohua Song, Hanping Hou, "On traditional M. F and Modern M. F", *Journal of Beijing Jiaotong University (Social Sciences Edition)*, Vol. 3, No. 1, pp10-16, 2004.

The Distribution Network Planning Using Geographic Information System (GIS)

Hai Yu¹, Linning Cai²

¹Department of Industrial Engineering, Tsinghua University, Beijing 100084, China

²Department of Industrial Engineering, Tsinghua University, Beijing 100084, China

ABSTRACT

To improve the customer service level, tobacco companies in China use the strategy of direct delivery. But this strategy has a great influence on the local company's operation cost. In this paper, a distribution planning of the local company is made to decrease the operation cost.

The distribution planning includes the location and choice of the distributed centers and the transferring points. In this problem, the Greedy Dropping Heuristic Algorithm is used twice to decide the distributed centers and transferring points. After the transferring points are set, the Vehicle Routing of each transferring points is presented. The savings algorithm is used in this process.

To analyze the complicated data, as well as to make the decisions, GIS (Geography Information System) technology is used, which turns to be a great help.

Key words: logistics, distribution, inventory, vehicle routing problem

1. INTRODUCTION

To improve the customer service level, tobacco companies in China use the strategy of direct delivery, delivering tobacco from local company directly to their customers, i.e., different retailers. But this strategy has a great influence on the local company's operation cost, which includes the labor cost of delivery workers, the cost of delivery vehicles and the inventory cost. All the cost mentioned above adding up will be a great number. For example, for a local company studied, the delivery cost counts up to about 7 million RMB, while the inventory cost is as much as 35 million RMB. So in this paper, a distribution planning of the local company will be made to decrease the operation cost.

The project is divided into two parts; the first is to determine how many distribution centers will be used. Now in their system, there are 42 distribution centers, including 11 bigger one belonging to 11 sale divisions and 31 smaller one dispersed in different towns in convenience for the distribution of tobacco. Considering the position and scale of all the 42 distribution centers, as well as the traffic conditions, several candidate schemes consisting of different distribution centers are put forward. After the first phase work, the main nodes of the distribution network are determined.

Based on the above structure, the second one is to determine the route planning for every demand point, i.e., the basic node in the network consisting of certain number of retailers. Considering the laws and the working load of the workers, the working time of worker is set to about 6 hours. The truckload is supposed to be big enough at first. Then using some heuristics of Vehicle Routing Problem, the route planning for every

demand point is decided. According to the total demand of all the demand points in the same route, the truck load of the vehicle and which kind of vehicle to use are decided at last.

To analyze the complicated data, GIS (Geography Information System) technology was used. The correct and ample geography information is needed in the system. So the preparation of the geography information is the first and most important step. Then all the nodes of the distribution network are presented in GIS, including the attributes of the nodes. Thus we can see clearly the distributing of the customers and the traffic conditions of them. It will be a great help for the decision that how many distribution centers should be used. When the routes computed out is displayed in GIS, we can see clearly which route is unreasonable in the scheme and thus correct it.

2. GIS TECHNOLOGY AND ITS USE IN LOGISTICS

In the beginning of the evolution of the vehicle routing systems, routes were designed manually with the aid of paper maps that provide the locations of the customers. From the early 1980s the development of computer technology gave birth to the systems for analyzing and displaying spatially or geographically related data. In the 1990s, the rapid growth of GIS gives us the hope of good use in logistics. However the most prevalent GIS software can only deal with and display the data that is formatted. While in the real life, the dimension of the location and routing problems are quite large. To design routing more efficiently, it is necessary to develop decision support systems that incorporate effective methods and spatial techniques.

Considering the present conditions of the local company, 28 towns that have warehouse of themselves are regarded as the candidate transferring points, which are presented in figure 2. Also the Greedy Dropping Heuristic Algorithm is used to derive the optimal choice of the transferring points, while a constraint is added: When the candidate transferring point is removed and the customers belong to it are re-distributed, if there exists more than one customer whose distance to the transferring point is more than one and a half hours' traveling time, the transferring point should not be removed.

As the result 18 transferring points are selected to transfer tobacco to the customers. Then the 18 transferring points are distributed to the 3 distribution centers, which are presented in figure 3.

Figure 3 The scheme of 3 distribution centers and 18 transferring points

5. THE VEHICLE ROUTING OF EACH TRANSFERRING POINTS

For each transferring point, the customers belong to it should be distributed from it and in the end return to it as the depot. In the real problem, the importance of choosing proper algorithm and modifying the routing according to the actual conditions are equal. In the solution of this problem, the Savings Algorithm presented by Clarke and Wright in 1964 is used. While in this algorithm, the truckload constraint and the time constraint are considered at the same time.

Then the results of the vehicle routing are presented in the geographical information system. With the aid of the experience of the workers and the local transportation

conditions etc., the routing schemes are modified. Figure 4 presents an example of the vehicle routing of one transferring points.



Figure 4 One example of the vehicle routing

6. CONCLUSIONS

From the above we can see that whenever the choice and location of the distribution centers or the choice of the transferring points or the vehicle routing of each transferring problem, the GIS technology is a great help in all the decisions. From complex and dull various data to the simple and sensed view of the map, the human can take part in the decisions directly.

Finally, the distribution plan comes to be successful, which decrease the local company's operation cost by 10% or so.

REFERENCES

- [1] Lawrence Bodin, Lawrence Levy, "Visualization in Vehicle Routing and Scheduling Problems", *ORSA Journal on Computing*, Vol. 6, No. 3, pp261-269, 1994
- [2] C.D. Tarantilis, C.T. Kiranoudis, "Using a spatial decision support system for solving the vehicle routing problem", *Information & management*, Vol. 39, No. 5, pp359-375, 2002

The Impact of IT Capability and E-Business on the Supply Chain Performance

Linglang Tang¹, Chiachi Shih¹, Chengchuang Hon²

¹ Graduate School of Management, Yuan-Ze University, Chung-Li, Taiwan, 320, China

² Minghsin University of Science and Technology, Hsin-Chu, Taiwan, China

¹ E-mail: balltang@saturn.yzu.edu.tw

ABSTRACT

This research has developed an analytical framework based on resource-based theory. It intends to find out the impact of different attributes on supply chain performance and financial performance under different IT capabilities and application degrees. The hypothesis test model incorporating five measuring variables is conducted; they are strategy of e-business, application degree of e-business, information technology capability, supply chain performance and financial performance. The results indicate that e-business strategy will affect the application degree for either high-tech industry or traditional industry; and higher IT capability will make the application degree impact on supply chain performance more significant while financial performance is not obviously activated. E-business should be a strategic policy. The managerial full support and involvement, cautious selection of new operational strategy, as well as the integration of e-business process and information, will make IT application, e-business process and readiness collaborate closely, so as to create business value and promote operational performance of supply chain.

Keywords: application degree of e-business, information technology capability, supply chain performance, resource-based theory

1. RESEARCH BACKGROUND AND MOTIVE

Coming along with the knowledge economy, the trend of globalization, digitization unceasingly challenges existed management concepts. The development of e-business has become the primary consideration to be survived. Information technology (IT) is applied to promote operational efficiency in the past. Nowadays, due to the maturity of internet technology, it has been applied to commercial management, and to integrate the information system. The development of e-business will undoubtedly induce further revolution of supply chain (SC). The construction of business to business (B to B) can especially improve the efficiency and circulation speed of information, enhance the effective management on supply chain, and decrease the risk and pressure of stock.

The rapid change of IT speeds up the footstep of e-business and the integration of business process across different enterprises. Upstream and downstream over the supply chain system can promptly and correctly link their logistic flow, information flow and cash flow together. The efficiency and effectiveness of whole supply chain is upgraded, and the demand uncertainty is decreased. Supply chain management can therefore be greatly improved. However, whether the competence of e-business can be improved depends on its ownership of flexible information technology. Only those who own superior information capability can realize the competitive advantage in the e-environment.

The application of information system has positive influence on supply chain management in many companies (Barua *et al.*, 2001). Most of the literatures focused on the theoretical study, and little of them

discussed in empirical way. This paper will measure the IT capability of enterprise by using resource-based theory. The impact of e-business on supply chain performance and financial performance will be analyzed under different IT levels and corporate features, in order to make concrete conclusion and suggestion. If the enterprise can integrate the inner and outer management information system while constructing e-business, the supply chain performance and competitive strength can be greatly improved.

The major objective of this paper is to investigate the impact of e-business on supply chain performance and financial performance for manufacturers under different IT levels and corporations.

2. EB STRATEGY AND IT CAPABILITY

2.1 E-business and Strategy

E-business is the generic name of business models and processes accomplished by internet technology, it focuses on the improvement of long term business performance (Simchi-Levi *et al.*, 2000). E-commerce corresponds to the ability to operate business in electronic way. So e-commerce is only one part of e-business; internet is the driving force of corporate re-engineering. E-business emphasizes more on the development of extended business, that is, business interior, business to customer (B to C) as well as business to business (B to B) transactions.

The applications of e-business onto supply chain have many different configurations. Lee *et al.* (2002) divided e-business applications into e-commerce, e-procurement and e-collaboration.

- (1) E-commerce: Assisting the supply chain partners inside the network system can quickly confirm or feedback via internet to fulfill customer’s requirement.
- (2) E-procurement: Corporations can access direct or indirect materials, handle value-added services such as transportation, warehousing, payment mechanism, quality assurance and documentation management etc. via on-line mechanism.
- (3) E-collaboration: Accelerate the coordination and cooperation of various transaction activities among the supply chain partners including all suppliers and customers exploiting the internet.

Johnson *et al.* (2002) defined e-business as the marriage between the integration of internet and supply chain. This marriage transformed many processes within the supply chain from procurement to customer management and product design. The enterprise must integrate the information system with their developed e-business capability. The effective integration can systematically utilize all resources and take advantage of leverage function. Grandon and Pearson (2003) identified three factors to be influential concerned with the perception of strategic value in other information technologies: operational support, managerial productivity, and strategic decision aid.

Barua *et al.* (2001) suggested a e-business value model which included e-business drivers, operational excellence measures and financial measures as measuring indexes. E-business drivers encompassed three areas: e-business processes for customers and suppliers; IT applications for customers, suppliers and internal operations (all of them must be integrated); and the e-business readiness of customers and suppliers. They stated that in trying to bring about e-business transformation, companies generally focused too much of their attention on technology. But systems didn’t work in a vacuum, and senior managers must recognize the complementary nature of technology, business processes and e-business readiness throughout the value chain.

2.2 IT capability

Mata *et al.* (1995) stated that managerial IT skills were rare and firm-specific which might be the source of sustained competitive advantage. The resource-based view assumed firms possessed four kinds of resources that differentiated it from other firms; that is, value, rareness, imperfect imitability and insubstitutability (Barney, 1991). The dynamic capabilities were defined as the firm’s ability to integrate, build and reconfigure internal and external competences to create profit and achieve competitive strength (Teece, Rumelt, Dosi and Winter, 1994; Teece, Pisano and Shuen, 1997).

Broadbent and Weill (1997) suggested IT infrastructure should include technical IT infrastructure and human IT infrastructure. Broadbent, Weill and St. Clair (1999)

pointed out the organization should be equipped with physical IT infrastructure such as hardware, software, communication technology, data base, application etc., as well as human IT infrastructure such as technique, experience, ability and knowledge etc. Via the combination of technology and human, the shared IT service can be constructed to provide services, such as networking, data access and exchange, business process support etc., to decrease operational cost and improve the performance and competence of organization.

Byrd & Turner (2000) summarized IT capabilities into eight dimensions, they are IT connectivity, applications functionality, IT compatibility, data transparency, technology management, business knowledge, management knowledge and technical knowledge.

As for the supply chain performance, Frohlich (2002) indicated that operational performance was based on faster delivery times, reduced transaction costs, and enhanced inventory turnover related to web-based integration. Operational excellence measures included the percentage of online business, the percentage of online procurement, the percentage of online customer-service provided online and order delivery time. Financial measures included revenue per employee, gross profit margin and return on assets.

3. DESIGN QUESTIONNAIRE SURVEY

3.1 Survey Structure

The structure of this survey includes five factors as illustrated in Figure 1. IT capability and corporate feature are treated as moderating variables in order to study whether they will affect the relationship between e-business and SC performance. The relationship between hypotheses and various variables will be validated empirically in this paper.

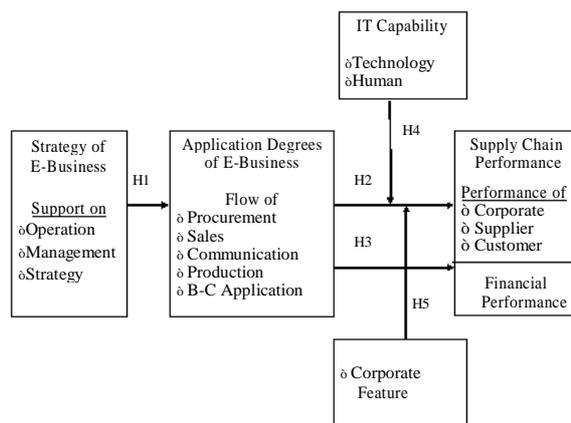


Figure 1 Research structure

Synthesizing the relevant literatures, we generate five hypotheses:

H1: Application degree of e-business is positively related to strategy (goal) of e-business.

H2: Supply chain performance is positively related to application degree of e-business.

H2a: Operational performance is positively related to application degree of e-business.

H2b: Supplier performance is positively related to application degree of e-business.

H2c: Customer (buyer) performance is positively related to application degree of e-business.

H3: Financial performance is positively related to application degree of e-business.

H3a: Gross profit margin is positively related to application degree of e-business.

H3b: Net profit margin is positively related to application degree of e-business.

H4: The impact of application degree of e-business on supply chain performance is significant dependent on IT capability.

H4a: The impact of application degree of e-business on operational performance is significant dependent on IT capability.

H4b: The impact of application degree of e-business on supplier performance is significant dependent on IT capability.

H4c: The impact of application degree of e-business on customer (buyer) performance is significant dependent on IT capability.

H5: The impact of application degree of e-business on supply chain performance is significant dependent on corporate feature.

H5a: The impact of application degree of e-business on supply chain performance is significant dependent on the industrial type.

H5b: The impact of application degree of e-business on supply chain performance is significant dependent on organization scale.

3.2 Questionnaire Design

There are five factors considered in our survey sheet, they are strategy of e-business, degree of e-business, information technology capability, supply chain performance, financial performance. Five factors contained in each characteristic are three, five, seven, twelve (4, 4, 4) and two items in respective. Survey data is measured in ordinal scale (1, 2, 3 ...) according to Likert's five grading scales (e.g. 1 represents "contribute no benefit", 5 represents "contribute full benefit"). The financial performance is mainly obtained from TEJ data base.

4. RESULT ANALYSIS AND DISCUSSION

4.1 Data Collection

The investigation survey sheet is designed originally on literature review and revised through experts interview, so that the contents effectiveness can be well assured. The survey population comes from Taiwan Top 1000 Manufactures, as well as the corporate joined AB E-Business Program (An e-business promotion program focusing on procurement and customer order processing that is supported by Taiwan, China). Survey data is randomly sampled and collected through e-mail and paper-mail, we issue 830 samples and 105 effective sheets (12.72%) are collected.

The received 105 effective investigation sheets covered different corporate features, with which can be used for analysis in effectiveness. The statistical software tool SPSS for Windows 10.0 is used for corresponding data analysis, it is also applied for variables validity and reliability assured.

Confidence analysis assures the consistency and stability of the measuring result. Confidence analysis takes the functionality to measure the degrees of data confidence within each factor and its related attributes; if the confidence Cronbach α value is greater than 0.7 then the investigated data is assumed to be sufficient confidence. In general, the Cronbach α value for each factor is greater than 0.8.

4.2 Declarative Statistics

The surveyed manufacturers are categorized into three groups (high-tech co. x 52; traditional co. x 43; foreign co. x 11). Declarative data analysis is applied based on 5 factors and its related attributes in correspondence to three groups of corporate organization, with which it can show the difference among them (Table 1).

In general, the foreign corporations (all of them are electronics industry) invest more in e-business and IT than local industries do. The strategies of e-business are mostly for the purpose of operational support. It implies most of the enterprises perceive e-business as the tool for customer service improvement, distribution channel development and competitiveness enhancement. The application degrees of e-business on procurement and sales flow are lower, because they involve not only the enterprise itself, but also the integration with upstream and downstream enterprises. The implementation is more difficult than other items that make their application degrees the lowest. Currently, the application of e-business is mainly focused on internal communication flow, such as using e-mail to transmit information. The IT capability of foreign electronic corporations is better than that of local high-tech corporations, while the later surpasses that of local traditional industries. As for the performance, information sharing and management perform better.

Table 1 The comparison among strategy, application degree, IT capability and SC performance

Main-factor	Sub-factor	High-tech industry	Traditional industry	Foreign corporation
Strategy	Operational support	4.109	4.092	4.364
	Managerial productivity	4.020	4.006	4.409
	Strategic decision support	3.993	3.983	4.242
Application degree	Procurement flow	2.439	2.607	3.299
	Sales flow	2.554	2.589	3.082
	Internal communication flow support	3.758	3.513	4.136
	Production process	3.119	2.875	4.068
	Business process of B to C	3.177	3.258	3.879
IT Capability	IT connectivity	3.621	3.613	4.136
	Application functionality	3.454	3.425	4.227
	IT compatibility	3.566	3.500	4.046
	Data transparency	3.347	3.331	3.750
	Business operation knowledge	3.648	3.575	3.925
	Management related knowledge	3.517	3.508	3.667
Corporate operational performance	Professional IT knowledge	3.862	3.546	3.950
	Order fulfillment management	3.728	3.872	4.000
	Information flow management	3.898	3.910	4.350
	Procurement management	3.687	3.932	4.000
	Inventory management	3.602	3.667	3.700
Supplier performance	Intensify information sharing	3.510	3.795	4.000
	Shorten delivery lead time	3.388	3.718	3.800
	Increase delivery reliability and service level	3.347	3.590	3.800
	Decrease delivery cost of supplier	3.204	3.590	3.400
Customer (buyer) performance	Intensify information sharing	3.551	3.641	3.700
	Improve customer service level	3.571	3.692	4.000
	Increase customer satisfaction	3.571	3.692	3.800
	Upgrade customer royalty	3.347	3.615	3.600

4.3 Correlation Analysis

Pearson analysis is used to review whether the paired comparative data is significant difference, eight attributes are considered in Pearson analysis. Parts of the results can be summarized as below.

- (1) Strategy and application degree of e-business have positive correlation ($r=0.461$; $p \leq 0.01$).
- (2) IT application degree and supply chain performance (corporate operational performance, supplier performance and buyer performance) have positive correlation ($r=0.538$).
- (3) IT capability and supply chain performance has positive correlation ($r=0.578$; $p \leq 0.001$).
- (4) Application degree and financial performance has no significant correlation.

4.4 Regression Model Analysis

- (1) The relationship among strategy, application degree and supply chain performance
The regression model is used further to validate the hypotheses and clarify their linear casual relationships. Table 2 shows that the strategy will affect the application degree of e-business, that is, the more the

enterprise learn what benefit e-business will bring for them, the eager they will construct the operation management model of e-business. Hypothesis 1 is therefore supported by this result. The higher application degree will induce higher supply chain performance that supports Hypothesis 2. Based on the correlation analysis and regression model, the application degree and financial performance has no significant relationship, that is, Hypothesis 3 is invalid. The construction of e-environment and introduction of IT require enormous investment. Successful implementation of e-business also needs the integrated support coming from technology, process and human. The operational performance can be improved in short term, but the financial performance needs long term efforts and the integration of systems.

- (2) The impact of application degree on supply chain performance under different IT capabilities
Based on Hypothesis 4: IT capability will affect the relationship between application degree and supply chain performance, we sort surveyed corporations into two groups (higher IT capability and lower IT capability) to analyze the impact of application degree on supply chain performance. Hypothesis 4 is supported based on the analysis of Table 3.

Table 2 The regression model among strategy, application degree and supply chain performance

Dependant variable	Prediction variable	R ²	β	t value	p value	Hypothesis
Application degree	Strategy	0.212	0.461	5.142	0.000***	H1
Order fulfillment management	Application degree	0.260	0.510	5.803	0.000***	H2a-1
Information flow management	Application degree	0.189	0.435	4.729	0.000***	H2a-2
Procurement management	Application degree	0.206	0.454	4.990	0.000***	H2a-3
Inventory management	Application degree	0.174	0.417	4.501	0.000***	H2a-4
Supplier performance	Application degree	0.359	0.599	7.328	0.000***	H2b
Customer (buyer) performance	Application degree	0.332	0.576	6.906	0.000***	H2c
Gross profit margin	Application degree	0.003	0.055	0.394	0.695	H3a
Net profit margin	Application degree	0.035	0.188	1.350	0.183	H3b

Remark: *** Significant level $p < 0.001$, ** Significant level $p < 0.01$, * Significant level $p < 0.05$

Table 3 The regression model between supply chain performance and application degree under different IT capability

Dependant variable	IT capability	Prediction variable	R ²	β	t value	p value
Operational performance	High	Application degree	0.107	0.327	2.423	0.019*
	Low		0.095	0.308	2.173	0.035*
Supplier performance	High	Application degree	0.274	0.523	4.297	0.000***
	Low		0.162	0.403	2.954	0.005**
Buyer performance	High	Application degree	0.323	0.568	4.836	0.000***
	Low		0.109	0.330	2.341	0.024*

Remark: *** Significant level $p < 0.001$, ** Significant level $p < 0.01$, * Significant level $p < 0.05$

(3) The impact of application degree on supply chain performance under different corporate features

Hypothesis 5a proposes that the industrial type will affect the impact of application degree on supply chain performance. Based on the linear regression model, we observe that application degree has significant impact on supply chain performance, and IT capability will affect the supply chain performance no matter what industry group the corporation is belonging to. This means that H5a is not accepted.

Hypothesis 5b suggests that the organization scale will affect the impact of application degree on supply chain performance. The organization scale is divided into two groups based on the employee population – 1 stands for the population is greater than the mean value while 2 smaller than it. The result shows that application degree has significant correlation with supply chain performance in both groups, which means H5b is not accepted too.

4.5 Discussion

The validation of the hypotheses proves that all hypotheses are true except H3 (the financial performance and application degrees of e-business have no significant correlation.) and H5 (the impact of application degree of e-business on supply chain performance is not significant dependent on corporate feature.)

5. CONCLUSION

This paper investigates the relationship among strategy of e-business, application degree of e-business, IT capability, corporate feature, supply chain performance and financial performance. The results obtained from the analysis can be summarized as below:

(1) Application degree of e-business is positively related to strategy of e-business.

There is no doubt that implementation of E-Business will increase the competitiveness, save the cost and reduce the risk. However, the implementation of E-Business involves the strategic consideration of the business process. Top management must clearly figure out its process re-engineering project, set unique operation model and definite target for whole supply chain system, so that the competitive model of e-business can be constructed.

(2) Supply performance is positively related to application degree of e-business.

The higher application degree of e-business will bring about better supply chain performance. E-business is becoming the key success factor of competitive strength in the global environment. The brand new operation model created by e-business can shorten the unforeseen distance with partner, and control the stock status of upstream and downstream firms whenever it is necessary. The immediate, flexible and speedy operation ability earns the competitive strength for enterprise in the global fast changing environment.

(3) Financial performance is not significantly related to application degree of e-business.

The promotion of e-business should be sustained to improve the operational performance of business, until then, the profit margin can be increased further. Besides, the introduction of e-business and related information system require vast investment. Only long term continuous implementation of e-business can actually improve the financial performance.

(4) The impact of application degree on supply chain performance is significantly dependent on IT capability level.

E-business is the utmost application of IT. While applying e-business to business model and flow, a company must integrate it with its information system. Only the successful integration can make all resources emerge their multiplicative effect.

(5) The impact of application degree of e-business on supply chain performance is significantly dependent on industrial type.

A company should be capable to adapt itself to the dynamic fast changing environment. In general, the high-tech industry has capital-intensive and flexible information system, so the implementation of e-business is easier than traditional industry. But the encouragement from government has also made many large-scale traditional industries introduce e-business to improve their supply chain performance.

Most information industry in Taiwan is OEM or ODM factory. The short life time of product, the dispersion of suppliers, the fierce competition in the market as well as many uncertain factors might disturb the supply chain performance while promoting e-business. The market uncertainty in tradition industry is much less than high-tech industry, so the successful implementation of e-business will greatly improve its supply chain performance.

(6) The impact of application degree of e-business on supply chain performance is significantly dependent on organization scale.

E-business implementation is long-term strategic change, hence, the acceptance coming from employees plays an important role in the change of process. Top management must be well aware of e-business implementation and steer the process change with suitable promotion methods based on employees' traits. The operational efficiency of business process can therefore be improved.

REFERENCES

- [1] Barney, J.B., "Firm Resources and Sustained Competitive Advantage", *Journal of Management*, 17(1): 99-120, 1991.
- [2] Barua, Anitesh, Prabhudev Konana, Andrew B Whinston and Fang Yin, "Driving E-business Excellence", *MIT Sloan Management Review*, 43(1): 36-44, Fall 2001.
- [3] Broadbent, M., and P. Weill, "Management by Maxim: How Business and IT Managers Can Create IT Infrastructures", *Sloan Management Review*, 38(3): 77-92, 1997.
- [4] Broadbent, M., P. Weill and D. St. Clair, "The Implications of Information Technology Infrastructure for Business Process Redesign", *MIS Quarterly*, 23(2): 159-182, 1999.
- [5] Byrd, T. A. and D.E. Turner, "Measuring the Flexibility of Information Technology Infrastructure: Exploratory Analysis of a Construct", *Journal of Management Information Systems*, 17(1): 167-208, Summer 2000.
- [6] Frohlich, Markham T., "E-integration in the Supply Chain: Barriers and Performance", *Decision Sciences*, 33(4): 537-556, Fall 2002.
- [7] Grandon, Elizabeth and John M. Pearson, "Strategic Value and Adoption of Electronic Commerce: An Empirical Study of Chilean Small and Medium Business", *Journal of Global Information Technology Management*, 6(3): 22-43, 2003.
- [8] Johnson, M Eric; Seungjin Whang, "E-business and Supply Chain Management: An Overview and Framework", *Production and Operations Management*, 11(4): 413-423, 2002.
- [9] Lee, H. L., P. Padmanabhan, and S. Whang, "Supply Chain Integration over the Internet", *Supply Chain Management: Models, Applications, and Research Directions*, 2002.
- [10] Mata, Francisco J., William L. Fuerst and Jay B. Barney, "Information Technology and Sustained Competitive Advantage: A Resources-Based Analysis", *MIS Quarterly*, 19(4): 487-505, Dec 1995.
- [11] Simchi-Lei, D., P. Kaminsky and E. Simchi-Levi, *Designing and Managing the Supply Chain – Concepts, Strategies, and Case Studies*, McGraw Hill, 2000.
- [12] Teece, D.J., R. Rumelt, G. Dosi and S. Winter, "Understanding Corporate Coherence: Theory and Evidence", *Journal of Economic Behavior and Organization*, 23(1): 1-30, 1994.
- [13] Teece, D.J., G. Pisano and A. Shuen, "Dynamic Capabilities and Strategic Management", *Strategic Management Journal*, 18(7): 509-533, 1997.

The Implementation of an Intelligent Logistics Tracking System Utilizing RFID

Chichun Lo¹, Wenchuan Hsieh², Litsung Huang³

¹ Institute of Information Management, National ChiaoTung University, HsinChu City, Taiwan, China

^{2,3} Dept. of Information Management, ShuTe University, Kaohsiung County, Taiwan, China

¹cclo@nctu.edu.tw, ^{2,3}{wch, s90101160}@mail.stu.edu.tw

ABSTRACT

Due to globalization and the advancement of technology, the differences are lessened among businesses in terms of IT capabilities and product quality. Operation cost has become a critical key factor of success that determines company's profit level, as well as its competence in the market. Under the environment of electronic-commerce, logistics in the supply chain ensures a stable supply of stocks to satisfy customer's demands. An efficient supply chain requires a sound inventory system and operates in coordination with well designed shipping processes. Present shipping process uses barcode in tracking the flow of materials; however, scanning barcode one at a time is tedious and inefficient. Recently, RFID has been regarded as an important technology that influences the growth of global industries in the near future. RFID is able to effectively reduce the overall cost of logistics. However, application of RFID on supply chain management is still at its infant stage. Therefore, this research attempts to develop a prototype of a Smart Logistics Tracking System (SLTS) utilizing RFID. The proposed system aims to replace barcode with RFID, eradicate the time-consuming scanning procedure and ensures an efficient flow of materials. SLTS provides a cargo tracking platform that reports real-time location of the cargo and its flow.

Keywords: Radio Frequency Identification (RFID), Supply Chain Management (SCM), Electronic product code (EPC)

1. INTRODUCTION

With the ever greater computation power and affordable network communication, businesses incorporate Internet to automate order processing, purchasing, manufacturing, shipping, marketing, and various services through uniformed information exchange under predefined policies. Industry standard allows businesses to collaborate with up-and-down-stream partners. Supply-chain collaboration improves manufacturing efficiency, reduces holding and distribution cost, shortens time to respond to market change and enhances industry's competency as well as its viability. However, the development of logistics-associated technology is laggard compares to that of the information exchange. Present shipping process uses barcode in tracking the flow of materials; however, scanning barcode one at a time is tedious and inefficient. During the second US-Iraq War, US military teamed up SAVI Technology in building a real-time RFID logistics tracking system for military cargo. The use of RFID technology allows multiple ID scanning which accelerates the speed of material flow. Scholars and industries are now under research of making RFID an affordable technology for commercial use. The success of this attempt is promised to bring about a revolutionary progress in logistics and resolves various common inefficiency problems stemmed from bullwhip effect in supply-chain. In short, by utilizing RFID, businesses are able to deliver material efficiently and maintain a low holding cost. Majority of the present cargo transportation undergoes rather slow processes due to the use of barcode which hinders the speed of material flow. With RFID, the overall performance of supply-chain can be improved significantly. However, application of RFID on supply

chain management is still at its infant stage. Therefore, this research attempts to develop a prototype of a Smart Logistics Tracking System (SLTS) utilizing RFID. With SLTS, every unit of product is attached with an EPC (Electronic Product Code) tag before leaving the factory. From that point on, each handler of the down stream supply-chain (including those involved in the processes from factory to warehouse and warehouse to retailers until products are placed on the shelf) tracks the flow of goods by scanning RFID. In short, the proposed system is able to make distribution smoother the by integrating logistics information of the up-and-down-stream businesses.

2. LITERATURE REVIEW

2.1 Supply Chain Management (SCM)

According to Ccecil Bozarth[1], SCM is the active management of supply chain activities to maximize customer value and achieve a sustainable competitive advantage. It is a discipline that devotes in developing and running supply chains in the most effective and efficient ways possible. As shown in figure 1, supply chain activities cover everything from product development, sourcing, production, and logistics, as well as the information systems needed to coordinate these activities. There are four objectives of SCM[5]:

- 1 Consumer: SCM benefits the consumers by maintaining a stable supply of goods to keep the price low.
- 1 Distribution Center: SCM serves regional market, integrates all end points and handles mass distribution operation. Distribution center utilizes computer-aided equipments in processing

customer orders, managing inventory, shipping, controlling cost and etc.

- 1 Manufacturer: SCM enables quicker respond to demands and market changes. It offers flexible production plans, quality management and various logistics support. Overall, SCM enhances order processing.
- 1 Supplier: SCM cuts down inventory which leads to reduced storage cost. SCM prevents stocking unsoldable goods due to bullwhip effect or unexpected market change.

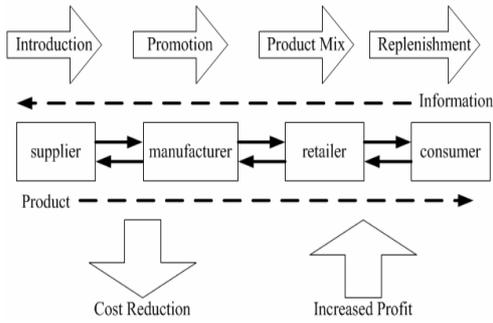


Figure 1. Supply Chain [6]

2.2 Current Supply Chain

The fundamental structure of present business organization is divided into departments including production, manufacturing, marketing, human resource, accounting and finance. The objective of all businesses are pursuing is to generate profits by selling products. Increasing incomes while reducing costs is the basic concept of survival.

Under current logistics system, factory delivers products to up-and-down-stream partners through supply chain within which routine inventory checks are applied at each stage. Businesses track goods by scanning the barcode attached on each unit using infra red reader. However, this process is rather inefficient because the items can only be scanned one at a time.

Therefore, this research attempts to develop a prototype of a Smart Logistics Tracking System (SLTS) utilizing RFID. The proposed system aims to replace barcode with RFID, eradicate the time-consuming scanning procedure and ensures an efficient flow of materials. SLTS provides a cargo tracking platform that reports real-time location of the cargo and its flow.

2.3 Radio Frequency Identification (RFID)

RFID is an identity management technique that utilized radio frequency (RF) to transfer and intercept identity information. RFID couples radio frequency identification technology with highly miniaturized computers that enable products to be identified and tracked at any point along the supply chain [2]. The system could be applied to almost any physical item, from ballpoint pens to toothpaste, which would carry their own unique information in the form of an embedded chip [3]. The chip sends out an identification signal allowing it to communicate with reader devices and other products embedded with similar chips [4].

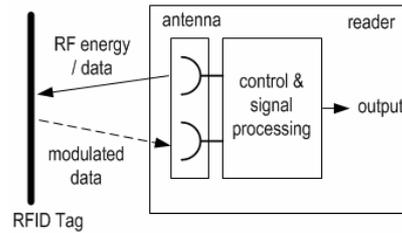


Figure 2. RFID structure [17]

A typical RFID structure (Fig 2) contains two major components: RFID tags and a reader. RFID Tag (transponder) contains antenna and memory. The antenna receives and responds queries from RFID reader, while memory stores ID information. There are two types of RFID tag: passive and active. Majority of RFID tags in existence are passive which converts RF energy into electrical power to generate respond signal. On the other hand, active tags require external power from embedded battery to generate initial signal. RFID reader (interrogator) is consisted of antenna, signal processing module and control circuits. When object identification is required, reader issues query signal. Those tags within communication range receive the query will automatically respond with its ID. The reader intercepts the respond signals and retrieves the IDs that identify the items, their locations or statuses.

Currently, the concept of RFID still requires an integrated and systematic mechanism, including data content and technical / software standards. In comparison with other identification technologies, RFID has the following advantages:

- 1 Faster data retrieval
- 1 Multiple tag read/write access
- 1 It is wireless (long read range)
- 1 No line-of-sight requirement
- 1 It is resistant to harsh condition
- 1 It is reusable (reduce cost)
- 1 RF is capable of penetrating paper, wood, plastic or other non-metal materials

Table 1. comparison between various systems

	bar code	optical	vocal	finger print	smart card	RFID
storage density	low	low	high	high	very high	very high
	limited	easy	easy	hard	not possible	not possible
cost	low	mid	high	high	low	mid
access speed	slow	slow	very slow	very slow	slow	very fast
max access range	0-50 cm	<1 cm	0-50 cm	direct	direct	0-500 Cm

Table 1 shows the comparison between various identification systems. RFID stores data digitally hence is able to store more information. Also, RFID is securer

since data content is concealed within the chip. Besides, RFID provides faster ID access and longer access range. RFID is able to remotely identify personnel, live-stocks or objects that has RFID tags attached. RFID tags are commonly used for animal tracking, key/lock, anti-theft system and material tracking in the supply chain. As a matter of fact, there are enormous possible applications and potential business opportunities that should not be underestimated.

Electronic Product Code (EPC) is developed by Auto-ID center of MIT. EPC is a coding scheme, which is 96 bits long and can be segmented into four pieces: header, manufacturer, product and serial number (fig 3).

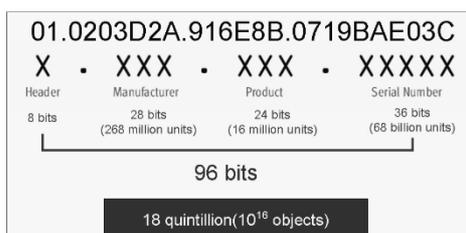


Figure 3. Electronic Product Code [20]

EPC supports current EAN standard manufacturer and product code which reduces the task of recoding [22]. EPC provides an encoding and decoding standard for RFID. A unique EPC-format identification data is written on each of the RFID tags. EPC will be requested upon item tracking and the data retrieved is then transferred to a shared database that serves interconnected companies [19].

2.5 Physical Markup Language (PML)

As defined in the PML Core specification [14], the goal of the Physical Markup Language (PML) is to provide a collection of common, standardized vocabularies to represent and distribute information related to EPC Network enabled objects. The PML vocabularies provide the XML definitions of the data exchanged between components in the EPC Network system. In other words, PML is used to describe EPC under network environment (fig. 4) and to enable internal and external communication among interconnected companies.



Figure 4: PML schema [10]

PML schema Core is comprised of the following elements:

- 1 Sensor Element: The Sensor element is the main interchange element for PML Core messaging. This element is a composite element comprised of the following subordinate elements:
 - n ID element
 - n one or more Observation elements
- 1 Observation Element: It contains the result of a measurement by a particular sensor. Each observation must be labeled with date and time. It can also be equipped with a unique ID, and a reference to the kind of command that was issued to make the observation. The Observation element consists of the following:
 - n an optional ID element
 - n an optional Command element
 - n DateTime element
 - n zero or more Data elements
 - n zero or more Tag elements
- 1 Data Element: The Data element must be used to represent the data captured when a sensor measured a particular property or entity, unless the data captured can be represented as Tag elements. The Data element consists of a choice of the following 3 elements:
 - n Text element: must be used to represent captured data as string notation.
 - n Binary element: must be expressed in hex.
 - n XML element: must be in XML format.
- 1 Tag Element: The tag element is a special kind of observed value introduced with recognition of the importance of automatic identifications in the EPC network. The “tag” entity represents any device that can be detected by a sensor. The Tag element consists of the following elements
 - n ID element
 - n optional Data element
 - n zero or more Sensor elements
- 1 ID Element: The ID element is used to capture tag identification information. The ID element has the following attributes.
 - n an optional schemeID attribute
 - n an optional schemeAgencyID attribute
 - n an optional schemeVersionID attribute
 - n an optional schemeURI attribute

2.6 Savant

Savant is “middleware” software designed to process the streams of tag or sensor data. Savant performs filtering, aggregation, and counting of tag data, reducing the volume of data prior to sending to Enterprise Applications [21]. The ONS provides a global lookup service to translate an EPC into one or more Internet addresses where further information on the object may be found. The characteristics of Savant are as follows:

- 1 Savant is the basic working platform of Auto-ID System.
- 1 It handles large amount of data.
- 1 It provides flexible data filtering and monitoring.
- 1 It controls the readers.
- 1 It provides interface between readers and application software or ERP.

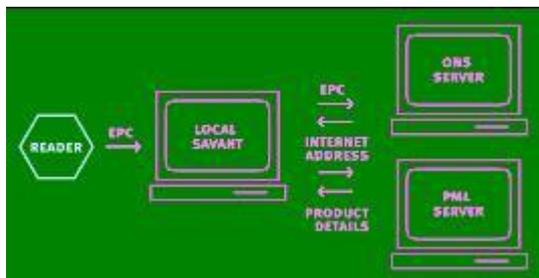


Figure 5. Savant Platform [21]

3. SYSTEM ARCHITECTURE AND DESIGN

3.1 System Framework

This research attempts to develop a “Smart Logistics Tracking System”(SLTS) utilizing RFID. SLTS consists of two major components: Logistics Tracking Center (LTC) and on-board Positioning and Tracking Module (PTM). LTC consists of ONS server and Savant; it is the central management panel that allows manager to monitor the material flow. PTM is a mobile tracking system that reports current shipping status to LTC.

Under proposed system, the traditional bar code scanning procedure is replaced with RFID technology which allows remote access of multiple data and has no line-of-sight requirement. Figure 6 shows the system framework and flow. Whenever item checking is required, RFID reader obtains IDs from tags, within its read range, then stores data to backend database. SLTS reduces resources and time required in the process.

As for logistics tracking, SLTS incorporates GPS (Global Positioning System) and GRPS (General Packet Radio Service) to achieve mobility and flexibility. GPS provides current location of the shipping items while GPRS serves as a communication channel for transmitting data. The mobile nature of SLTS enables flexible management and immediate response to changes; such capabilities are what barcode failed to provide.

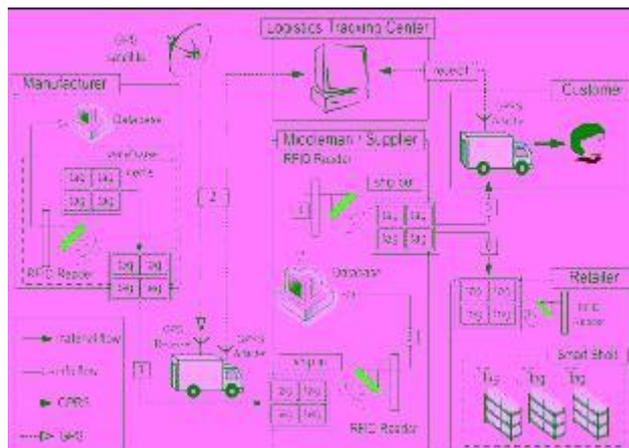


Figure 6. System Framework and Flow

Goods manufactured were attached with tags and stored in the warehouse. Before shipping, RFID reader examines the items delivered and updates the associated records in the database (step 1). If item tracking is requested while shipping, the GPS and RFID reader will report its current position and item information through GPRS to Logistics Tracking Center (LTC) (step 2). Upon arriving the supplier, the items were checked using RFID reader and database is updated accordingly (step 3). When shipping out the items, supplier uses RFID reader to perform item check and updates item information to database (step 4). Finally, as customer receives the products, RFID reader checks the item and the receipt will be confirmed with LTC through GPRS (step 5). Similarly, when receiving the items, retailer uses RFID to perform item check while confirming with LTC through GPRS. The items are then placed on the RFID-enabled smart shelves (step 6).

3.2 Logistics Tracking and Positioning

The most prominent feature about SLTS is its ability of tracking and positioning. Figure 7 illustrates a sample instance of material tracking in the LTC. SLTS provides interface to adjust the zooming scale, allowing managers to remotely inspect the shipping items along with neighboring street details. The map on the left shows each destination with a transaction ID (TID). With TID the items to be delivered can be referenced.



Figure 7. Real-time tracking

Figure 8 shows the on-board positioning and tracking module (PTM), which consists of GPRS adapter, GPS

receiver and embedded micro processor. The processor is responsible of coordinating GPRS, GPS and RFID reader, as well as gathering information to be transferred back to the LTC.

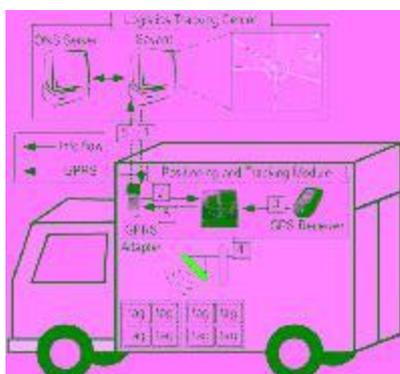


Figure 8. On-board Positioning and Tracking Module (PTM)

When item tracking is requested, LTC sends the query through GPRS to the target truck that loads the items (step 1). As soon as PTM receives the request, the micro processor commands GPS to reports it current position (step 2); at the same time, RFID reader scans the tags to obtain EPCs (step 3) for processor to verify the items being queried. The tracking result will be composed by processor and transferred through GPRS adaptor (step 4). Finally, LTC is able to reference more information by searching ONS based on the retrieved EPCs (step 5).

4. SYSTEM DEMONSTRATION

4.1 Logistics Tracking Center

Figure 9 shows the LTC's main system interface. The proposed LTC utilized MapExtreme to display the geographical information and maps out the current track locations. Section A is displays messages to inform user about the system's current status. Section B is the event list containing records that describe events reported by the monitor tracks. Each record is composed of several fields: event sequence, event name, event type, description and time.

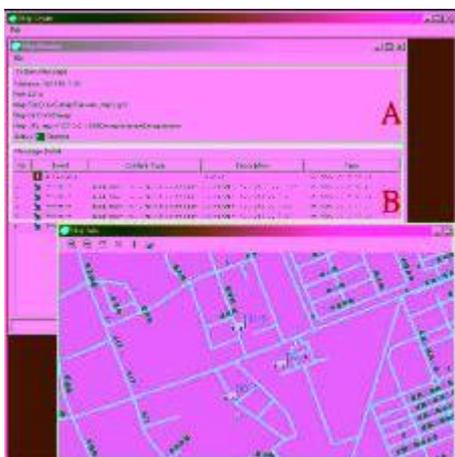


Figure 9. LTC main interface

4.2 Positioning and Tracking Module

Figure 10 shows the main interface of PTM which is implemented on a PDA running embedded Linux. Section A displays system status while section B keeps a list of even records indicating event, description and time occurred. PTM takes the coordination reported by GPS and send the information as in MMS format through GPRS back to the LTC.



Figure 10. Positioning and Tracking Module

5. CONCLUSION

A sound logistics system can improve businesses' flexibility and performance while reducing overall cost of shipping and storage. Business should incorporate information technologies into supply chain to achieve efficient data exchange and information sharing and effectively integrate up-and-down stream partners. The proposed SLTS is a system that utilizes RFID to improve the efficiency of the traditional supply chain. With the assistance of real-time monitoring facility, managers are able to keep track of material. However, further improvement should consider providing on-line shipment modification to offer greater flexibility to respond to the market change.

REFERENCES

- [1] Ccecil Bozarth, What is SCM, <http://src.ncsu.edu/public/s1/whatis.html>
- [2] Greg Jacobson, "Technology revolution underway." Chain Drug Review, October 22, 2001. Available online at http://www.chaindrugreview.com/articles/tech_revolution.html
- [3] Auto Center Joins UK Group, MIT TECH TALK, (Jan. 24, 2001), available at <http://web.mit.edu/newsoffice/tt/2001/jan24/auto.html>
- [4] Introduction to Auto-ID, available at <http://www.autoidcenter.org/technology.asp>
- [5] RFID home page, The Association for Automatic Identification and Data Capture Technologies, <http://www.aimglobal.org/technologies/rfid/>
- [6] A Low Power RF ID Transponder, By Raymond Page, Wenzel Associates
- [7] Peter Hawakes, "RFID Tag System", BTG Ltd, 1998

- [8] AIM Global network,
http://www.aimglobal.org/technologies/rfid/what_is_rfid.htm
- [9] RFID journal,
<http://www.rfidjournal.com/index.html>
- [10] Klaus Finkenzeller, RFID HandBook, Wiley, 1999, UK.
- [11] MicroID 125khz RFID System Design Guide. Microchip Technonlogy Inc., USA, 1998
- [12] Texas Instruments, "Ti-RFID Web"
<Http://www.ti.com/tiris/default.htm>
- [13] Chen, S.C.Q.; Thomas, V. "Optimization of Inductive RFID Technology", Electronics and the Environment, 2001, Page (s) : 82-87
- [14] Christian Floerkemeier, Dipan Anarkat, Ted Osinski, Mark Harrison, "PML Core Specification 1.0", Auto-ID Center, 2003/9/15
- [15] Giesecke and Devrient GmbH., Germany, RFID HANDBOOK Radio-Frequency Identification Fundaentals and Applications, 1999.
- [16] Steve C. Q. Chen and Valerie Thomas, "Optimization of Inductive RFID Technology", IEEE Electronics and Environment Proceeding, pp. 82-87, 2001.
- [17] Youbok Lee, "RFID Coil Design " Application Note AN678 by Microchip Technology Inc.
- [18] R.Schneiderman, "RFID Tags Locate Growing Wireless Markets," Microwaves & RF Magazine, Feb. 1994
- [19] microID™ 13.56 MHz RFID System Design Guide, 1999 Microchip Technology Inc.
- [20] R. Page, "A Low Power RFID Transponder," RF Design. pp31-34, July 1993.
- [21] Auto-ID Object Name Service (ONS) specification 1.0,
<http://develop.autoidcenter.org/TR/2003/WD-ons-1.0-20030930.pdf>

The Operating Cost in a Tele-working Organization

Xiaojun Li, Pengzhu Zhang, Fanghsin Liu

Aetna School of Management, Shanghai Jiaotong University, Shanghai 200052, China
{xi_anlxj, pzzhang, fanghsinliu}@sjtu.edu.cn

ABSTRACT

In this paper, we develop a model of the operating cost in a teleworking organization and describe the components and the changes of the operating costs. Then we give the application of the model for the decision-making in an organization willing to adopt teleworking.

Keywords: Teleworking; Operating costs; Net present value (NPV) criterion

1. INTRODUCTION

Telework, considered as an innovative work organization form, has emerged thirty years ago. However, at present telework has been adopted by more and more companies as a tool to cut costs, attract and retain top personnel in fields with short labor supplies[1]. Until 2003, 19.8 million persons usually did work at home as part of their primary job according to the Bureau of Labor Statistics of the U.S. Department of Labor report[2]. In Britain, official government statistics show telework has doubled in the six years since its measurement was incorporated into the Labour Force Survey. Whilst growth in some years has hit as much as 20 per cent, for 2003 it has been a more average of around 12 per cent. The total has grown from 1.01 million in 1997 to 2.113 million people or just over 7.5% of the workforce.[3]. In China, the government and companies' appeal to telework is also becoming stronger and stronger.

However, the cost is one of the key factors when an organization decides whether to adopt teleworking. The past study about the cost of telework is integrated with the cost components less and uncompletely. For instance, just as the author (On Cost In Teleworking) pointed out, the model didn't consider learning costs explicitly. Moreover, the existing studies about costs in a teleworking organization mostly are passive studies.

In this paper, our major interest is to understand the components and changes in costs when an organization changes from traditional form to teleworking form. The objective is to provide a decision support tool about the operating costs in a teleworking organization.

2. TYPES OF TELEWORK

Telework is defined simply as employees who work anywhere except the main office and make use of the electronic connections, for example telephone lines, cellular, wireless circuits, internet access. Telework has developed various kinds of types in different countries and locations until now, including mobile working, home-based teleworking, satellite offices, and

neighborhood work centers [1].

And it is culture to mainly influence on various kinds of types of works producing. In America, the freedom and independence is regarded as the most important factors in American literature so home-based teleworking are satisfied with the general American life. But, in Europe quantity of the staffs who spend their working time at home are less than in America.

3. ORGANAZATION OPERATING

If a company intends to adopt telework for the improvement of productivity, efficiency and flexibility or any else, the first one what need to do is to ascertain the new culture and than implementation completely. During the implementation, some objectors become the producer of obstacles for reformation but they will follow the majority's opinions and join the party by more and more staffs supporting the reformation. So to select a suitable pilot project is vital for an organization that has the will to adopt teleworking.

A good pilot project should have 5 elements including people, business, time, scope, and thing as the following:

- People (staffs): enthusiastic participants (volunteers) but few
- Expected reaction: criticizing the business moderately
- Business: constructing the new reasonable standard.
- Time: a period of less than one year
- Scope: suitable work scope
- Thing: a nominal cost

When the majority in the organization accepts the new company's operation that leads to the new company culture, it is time to decide which parts of works are suitable for teleworking. The following traits should provide the possible criterions for selecting the suitable department to perform the teleworking .

- Activities requiring concentration intensely and creative thought

- Individual or small-scope review and analysis
- Work that requires low levels of physical collaboration
- High telephone and computer use

However we can think over the problem from another aspect. In every organization, a variety of tasks have to be carried out by employees. In a given situation the employees of the company have a number of tasks. Each

of these tasks can be described by a number of variables or dimensions[5]. Examples of such dimensions are listed in Table 1 (On Cost In Teleworking, p56).

In fact, we can try to identify these characteristics of a task to decide whether it is suitable for teleworking. A task having the left characteristics is more suitable than the others in the table 1.

Table 1. Examples of task dimension

Degree of routinization	High <.....> Low
Degree of structuredness	High <.....> Low
Degree of active action necessary	Low <.....> High
Complexity	Low <.....> High
Need for internal communication	Small <.....> Large
Need for external communication	Small <.....> Large
Demand for competence	Small <.....> Large
Coordination need	Small <.....> Large
Time demands	Small <.....> Large
Environment	Stable <.....> Dynamic
Demand on accessibility	Low <.....> High

4. THE MODEL OF OPERATING COST

In an organization, some of the company’s costs vary with output, while others remain unchanged as long as the company is producing any output at all. Moreover, when the variable costs increases continually until it almost equals to the fixed costs, the leader in the company begins to think over another model to decrease the variable costs or change it into the fixed costs such as building up the new location or buying new equipments. With different parts of the cost, we apply the different analysis methodology. We therefore divide total cost (TC) –the total economic cost of production—into two components:

Fixed cost (FC): A cost that does not vary with the

level of output.

Variable cost (VC): A cost that varies as output varies.

Then FC can be divided assets costs, management costs and information costs. VC is composed of marketing costs, traveling costs, training costs, production costs, and other costs. The costs considered are shown in Fig 1.

4.1 Fixed costs

When an organization changes from one form to another with a given work, the fixed cost will vary. To begin with the discussion, we subdivide the fixed cost into three types of costs:

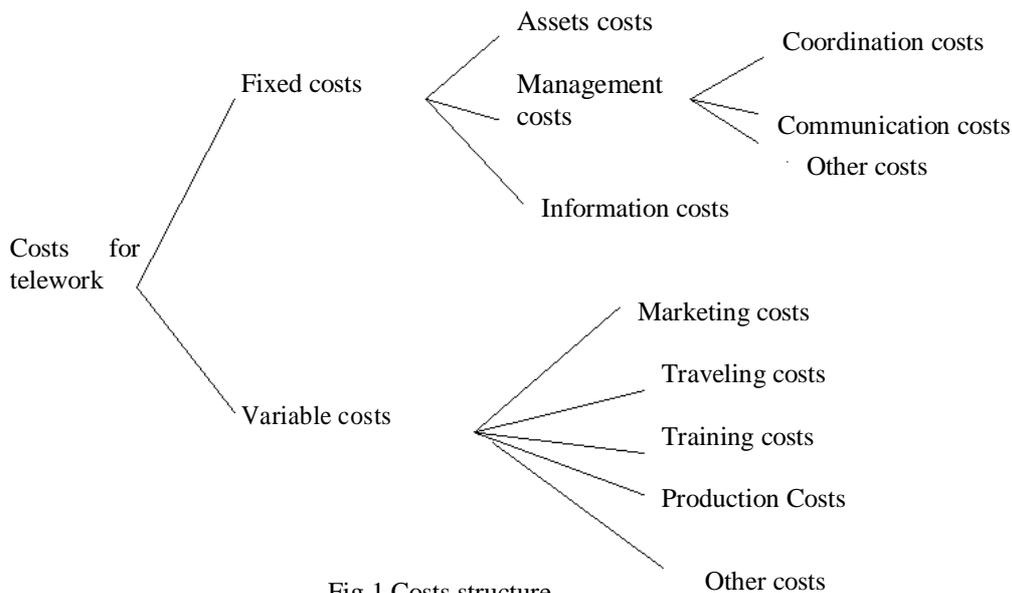


Fig 1 Costs structure

1. assets costs
2. management costs.
3. information costs

4.1.1 Assets costs

We refer to assets as the combination of working equipments and space.

Teleworker must establish “electronic connectivity” to the main office. Electronic information is typically conveyed in the form of voice, page, FAX, email or collaboration software application data. It can travel over a telephone line, cellular phone circuit, cable connection, or satellite link. Once the implementation mechanism for secure electronic connectivity is established, appropriate software packages should be procured to help mitigate the problems arising from reduced face-to-face communications. Applications could include uniform messaging (i.e., email, voice mail, and FAX all delivered to and sent from a web-accessible email address), voice-over-IP (i.e., internet telephony), teleconferencing, video conferencing, threaded discussion list management, workflow management, virtual whiteboards for concurrent document review and update, and document management (publish, check-in, check-out, version control). Organizations usually choose to standardize two mechanisms, in order to keep their information secure and their Information systems burden manageable:

1. The teleworker’s computing system capability and software application suite
2. The nature of the electronic connection, including telecommunications equipment

Once standards have been established, companies must select the degree to which they will underwrite the teleworker’s equipment and software. Some put the burden fully on the teleworker, some underwrite the required pieces, and some supply the teleworker with a pre-loaded, “approved” workstation. Factors that go into this decision include: data interoperability, teleworking budget, security, maintenance costs, and use as a recruiting/retention discriminator. To protect sensitive company information, teleworkers are frequently connected to their company’s intranet via a Virtual Private Network or a direct modem connection. Equipment expenses include (to the extent underwritten by the company):

- Computing equipment and software
- Office equipment (telephone, answering machine, cell phone, pager, copier, scanner, FAX)
- Office furniture (desk, chairs, file cabinets, bookcases, lights)
- Telecommunications (telephone, cellular phone, and cable connections; Internet access; Virtual Private Network)

The assets costs will go up as a result of buying new equipment. On the other hand, the working space cost will tend to be lower. When one of the departments of a

company move from an urban district to a rural one, the space cost will go down because of the lower price of the location and the local government’s grant.

Another problem we cannot ignore is the leaving unused of the assets. In the teleworking form, the leaving unused of the assets is inevitable sometimes. For example, when an employee chooses the home-based teleworking—three days working at home and the other days working at the main office, the leaving unused of the assets is serious. When we take opportunity cost into account, the assets cost will rise dramatically.

So, the assets cost can be expressed as

$$\text{Assetscosts} = \text{Equipmentcosts} + \text{spacecosts} + \text{opportunitycosts}$$

4.1.2 Management costs

Telework, in any form, has ramifications beyond simply changing the way or place in which an individual employee performs work: It can extend to remote managing and virtual teams. Remote managing occurs when managers are physically separated from their members who work away from the main traditional office and communicate with them by information technology. Virtual teams consist of team members who are geographically dispersed and who come together by way of teleworking technology. Each team member may be located in a traditional office setting, but the offices are not proximate to one another. Additionally, virtual team members may telework, such as a teleworker who is a member of a multiple-site team.

Thus, telework may have additional challenges that emerge when consideration extends beyond the isolated telework—both up (to managers) and across (to teammates) the traditional organization. These challenges mainly resulted from the following several inherent obstacles:

- Only certain kinds of work are appropriate to perform at a distance
- Teleworker socialization needs are difficult to satisfy
- Non-telecommuting co-workers may resent the teleworker
- Electronic communication fails to substitute for face to face contact perfectly
- Meetings become harder to schedule
- Emergency actions involving the telecommuter become harder to take
- Teleworkers may lose focus on the task at hand
- Managers have difficulty managing by results only

These challenges of management may be showed as:

- Performance monitoring
- Performance measurement
- Mentoring
- Synergy
- Managerial control

Organization culture

To overcome these challenges, managers have to devote more works than before. The management costs will rise following these works increasing. For example, the internal coordination costs will increase because the managers have to travel between the central office and the teleworking sites. When the managers give up these traveling and coordinate by information equipment, the costs will change in pattern but not in amount.

4.1.2.1 Communication costs

Communication form of telework is much more different from traditional form and also can makes us describe obviously the advantage and challenge in managing teleworking model so we discuss it more in this subsection. We give our attention to three parts of costs: the costs of social communication within the organization, the costs of searching for information for decision and the costs of handling remaining ambiguity. To improve the communication affection and get enough information for each teleworker, the organization has to use many IT equipments and devote much time into these activities.

On one hand, the e-flow of the information brings the communication a higher speed. On the other hand, due to the lack of the effective communications in the teleworking organization, feelings of isolation and an inability to complete work tasks fully become the main disadvantages of teleworking for the employee. All these factors will be the main obstacle to establish the organization cultures in a teleworking organization. Additionally, the employees may have difficulty in completing work assignments as they have less access to the necessary equipment (i.e., photocopier or fax machine). So the opportunity costs of the communication and will be much higher than before.

4.1.2.2 Coordination costs

The coordination costs include the costs of commanding—giving and taking orders for instructions or control—and the costs of project leadership. The cost of coordination largely depends on the way the relationship between the company and the dispersed unit is regulated and this, in turn, goes back to the types of tasks that have to be performed by the unit. Generally, this part of cost will be arising in a teleworking organization. For example, the internal coordination costs will increase mostly because the project leader has to travel between the rural and urban offices to give ‘face-to-face’ instructions.

The reason why we put the communication and coordination costs into the fixed costs instead of variable costs is that communication costs in teleworking vary little, almost stable during the period. In the traditional working form, it is easy to communication costs vary

with output. If the teleworking is adopted, the e-flow and IT equipment will make costs more stable compared with the traditional working.

4.1.3 Information costs

Information costs is shown on the marginal costs in the traditional model and it is covered into other costs such as the office’s rent. When teleworking is adopted, the information costs emerge clearly and the costs may decrease more and more by using IT equipments. Another reason of the decreasing of this part is that the distance between consumers and markets gets closer and closer in teleworking model. Here we give the definition of the information costs in telework that is produced when information is collected, transferred and switched by IT equipments for operating.

In addition, we must point out that the management and information costs have some superposition inevitably. For example, we cannot classify the cost when a manager exchanges information with a teleworker for daily work. However, when we refer to the information costs, we pay more attention on a teleworker rather than the manager.

4.2 Variable costs

Compared with the fixed costs, we mainly discuss the variable costs from the aspect of an individual employee. In other words, we mainly examine the average costs of every teleworker.

4.2.1 Marketing costs

About marketing costs, we must distinguish two kinds of conditions. First, if the marketing task is the project leaders’ job, such as in some consultant companies, the marketing costs with the client will keep stable due to the project leaders remain at the old location.

However, if the organization has the specific department to market with the client, the marketing costs may decrease especially when the purpose of organization’s executing teleworking is to reduce the distance between the company and the customers. In these cases, it is possible to locate part of the company close to an important supplier or customer, which could lower cost. Under these conditions, the satisfaction degree of the client will rise and the opportunity costs of marketing also go down.

4.2.2 Travelling costs

The most obviously benefit for the teleworker from teleworking is the reduction of the traveling costs.

Studies show that Teleworking Programs can contribute to 60% fewer driver miles, and traffic congestion is reduced as the number of employees physically traveling

to their places of employment decreases. It is estimated that each driver could save two hundred gallons of gasoline and four quarts of oil each year by utilizing alternative work places.^[7]

On the other hand, this travel-time reduction may result in over two hundred hours saved annually—time which can then be spent on family, hobbies, and other activities.^[7] So we can get that if an employee choose telework, his/her opportunity costs of the traveling will be much lower than before.

4.2.3 Training costs

In an organization, it is critical that both supervisors and employees feel they can share their concerns, and that they are capable of communicating their concerns cogently. However there are many difficulties in how to act as a qualified teleworker and how to manage teleworking when a traditional organization changes to telework. So it's necessary for managers and employees to improve their ability to make use of the electronic media for communicating effectively by training.

Supervisors of teleworkers provide important leadership, coordination, and team-building functions for their project teams. An introductory training session for telemanagers is recommended to:

- Learning the best practices of teleworker management
- Reviewing organizational telework policies and guidelines
- Understanding the needs of the teleworker
- Surface any concerns they may have about teleworking arrangements

Better-informed people make better-informed decisions. Introductory training for teleworkers helps inform them of:

- Organizational telework policies and guidelines
- Their communication responsibilities
- Best practices of telecommuting
- The challenges of to be a teleworker

Thus, for a teleworking organization, at the beginning of operating there must be a huge budget for the training costs. But when the managers and employees in the organization have been similar with teleworking, the training cost will be lower than before because these trainings can go on the teleworking web.

4.2.4 Production costs

The impacts of teleworking in the production costs including the following aspects:

Most readily apparent is an increase in employee productivity. While such results may be based partially on subjective assessments, many companies report productivity increases ranging from ten to forty percent.

Similar increases have been reported in positions such as data entry clerks, where gains may be objectively measured.^[8] While some may believe these gains result from employees working longer hours, a better explanation may be that teleworking (i.e., home) creates an environment more conducive to work. The teleworker is not subject to constant interruptions for quick questions or chats during the workday. Thus, the actual time spent working outside the formal workplace can produce a higher quality and quantity of work output. In addition to an increase in productivity, a decrease in absenteeism can be anticipated. Many absences occur due to an employee not being able to come to work for reasons including minor illnesses, emergencies, or personal appointments. When employees are able to perform some of their work as teleworker, absenteeism decreases, thereby minimizing the total costs.

The option of teleworking creates incentives for employees to remain with the company and thereby increases the retention rate and decreases recruiting costs. At the same time, teleworking allows the company to increase its base of potential employees, since the geographic location of available candidates greatly expands.

All what we have discussed as the above are explicit affections for a teleworking organization. However there are other advantages of teleworking existing as externalities for the company. Therefore, We should not be surprised when the government give a teleworking organization the grants. And the grants also affect the costs in teleworking just as what we point out in the next part.

4.3 Application of the model in decision

How should a company decide whether a task is suitable for teleworking with a given work? Suppose there is a task T for a company. If we will complete the task with teleworking, we need n employees and t unit times. The fixed costs are C_f , and the average costs of variable costs for every employee is C_v per one unit time. The government grants are denoted as G , and the expected profits of the task are π . R is the opportunity cost of capital. In other words, the correct value for R is the return rate that the company could earn on traditional condition.

Then we can use the net present value (NPV) criterion to make decision:

$$NPV = \frac{\pi}{(1+R)} - C_v * n * t - an(C_f - G) * t$$

where $an(C_f - G)$ is the annual cost associated with fixed cost minus the government grants.

The company should adopt teleworking only if that net

benefit is positive—i.e., only if $NPV > 0$.

5. CONCLUSION

In this paper, we divide the operating costs in teleworking into several parts. Then according to the different characteristics of every part, we adopt different method to analyze it and build the model. And we design a mathematical model to support making decision according to this model when an organization plans to adopt teleworking in it. However there are still many problems we cannot give an effective method to solve. For instance, when we come to the opportunity costs of capital, no one can give an accurate value about R . Under this condition we must ask for help with our experts. But, we are not certain with to which extent the estimated value is reliable when facing a new thing. All these suspending problems are interesting and challenges for us. Following with the development of teleworking in China, our society will pay more attention on this field.

ACKNOWLEDGEMENT

The paper is supported by NSFC Project 70271025 and 70121001.

REFERENCES

[1] Nancy B. Kurland, Diane E. Bailey, Telework: The

Advantages and challenges of Working Here, There, Anywhere, and Anytime, Organizational Dynamics, Autumn 1999, 53—68

[2] Work At Home In 2001,

<http://stats.bls.gov/news.release/homey.nr0.htm>, 2004-2-11

[3] Telework Statistics Increase Again,

<http://content.equaltelework.org/index.php/equaltelework/content/view/full/112/> 2004-2-11

[4] Hitoshi Mitomo, Toshiya Jitsuzumi, Impact of telecommuting on mass transit congestion: the Tokyo case, Telecommunications Policy 23 (1999) 741-751

[5] Jörgen Lindström, Birger Rapp, On Cost in Teleworking, International Transactions in Operational Research, vol.3, No 1, 1996, 51—63

[6] Huws U.Korte, W.B.&Robinson, S.(1990). Telework. Towards the Elusive Office. chichester:Wiley

[7] Telecommute America—Benefits, 1997 WL 15140080 [AT&T Survey], Oct. 21, 1997

[8] Debra Schepp & Brad Schepp, The Telecommuter's Handbook 7 (2d ed. 1995).

[9] Robert S. Pindyck, Daniel L. Rubinfeld, Microeconomics, fifth edition, Prentice Hall, 2001, 542—543

[10] Christos J. Georgiou, Petros S. Stefaneas, Strategies for Accelerating the Worldwide Adoption of E-Commerce, COMMUNICATIONS OF THE ACM, April 2002/vol.45. No.4ve,pp.145-151

The Virtual Service Value-Chain: Disruptive Technology Delivering Competitive Advantage for the Services Industry

John Hamilton

Director of E-Business, James Cook University, Cairns, Australia 4870
John.Hamilton@jcu.edu.au

ABSTRACT

Services have been investigated from a diverse range of approaches. Services encapsulate over 65% of global business, yet there are many gaps in the services knowledge base – particularly from an operations management perspective. This research investigates an emerging and truly disruptive business scenario – the service value-chain, from both a marketing and an operations approach. The service value-chain is defined as the flexible, dynamic, delivery of a service, or product, by a business's coordinated value chains (supply chains and demand chains working in harmony), such that a value-adding, specific, service solution is effectively, and efficiently, delivered to the individual customer.

The 'virtual service value-chain customer –business encounter model' is developed. Impediments to the development of a service value-chain are investigated to delineate future areas of business research. The virtual, information aspect of the services value chain customer 'touch-point' across the web interface offers a raft of new research possibilities and possible new pathways to competitive advantage. Eight key areas related to websites are offered as investigation areas. Business response mechanisms to the customer 'touch-point' must be focus on the competitive mix of strategic features.

Key Words: Services, Logistics, Service value chain, Strategic intelligence, Website, Customer

1. INTRODUCTION

The services industry provides services not goods[1]. This industry is moving towards globalization[2]. In 1870 the service sector employed slightly more than 20% of the U.S. workforce, whilst by 2002 it employed in approximately of 82% of the U.S. workforce, and 81% of the private sector GDP[3]. Services Management is a 'transfunctional' research area[4]. It covers areas including service quality[5], services encounters[6], and service execution[7]. Services operations management and services marketing provide other perspectives to services.

Definitions of service (and what constitutes a service) range from the narrow to the broad. In 1960 the Definitions Committee of the American Marketing Association (1960) defined services as: 'Activities, benefits, or satisfactions which are offered for sale, or are provided, in connection with the sale of goods'.

Examples of a service include: amusements, hotel service, electric service, transportation, the services of barber shops and beauty shops, repair and maintenance service, and the work of credit rating bureaus.

Judd[8] and Rathmell[9] promoted the service sector of the economy, and the true nature of services. Murdick, et al.[10] and Quinn, et al.[11] broadened services definitions to include all economic activities where output was not a physical product or construction, and was usually consumed when produced, and delivered as an intangible, added value to the customer - for example, travel comfort. Again services were redefined.

'Services are deeds, processes, and performances' [12], but they may also be tangible. Czinkota, et al.[13] split services into tangible areas involving: 1) people (fitness centres); or 2) possession processing (like freight transportation), and intangible areas involving: 3) mental stimulus like (education and religion); and 4) information processing (like banking, data processing). Often services are integrally enmeshed with manufactured goods, or to the delivering (or enabling) of goods. Thus the distinction between goods and services is imprecise, and no clear boundary between manufacturing and service firms exists[14]. According to Levitt[15], 'There is no such thing as a service industry. There are only industries whose service components are greater or less than those of other industries. Everybody is in service.' Taken to its logical conclusion, it could be suggested that all manufacturing is indeed a service, as it delivers something!

Encyclopedias including Britannica, Encarta, Columbia; and dictionaries including Oxford and Macquarie, offer a range of definitions of service. To meet these diversities we follow Rust & Metters[16] 'topologies' approach to group some of the complexities of services down onto models.

2. SERVICE TYPOLOGIES

From an operations and marketing perspective topology schemes for services have generally lacked empirically tested works. Empirical works [17],[18],[19], offer some key exceptions, but overall empirical services related research is in its infancy. A topologies approach, based on recent service industry models identifies key

knowledge gaps, and establishes possible empirical research areas.

3. SERVICE MODELS

Figure 1 presents Rust & Metters view of services. They grouped service models as customer models (external) or service provider models (internal).

Each model was then segregated, as shown in Figure 1, into two of the three models:

- 1) 'customer behaviour models' - dynamic models of customer retention (like loyalty) and stochastic models of customer behaviour (like satisfaction), and of customer behaviour (like churn rate or loss of a specific customer during a single service encounter);
- 2) 'service quality impact models' – aggregate models (like customer satisfaction effects) and disaggregated models (like financial impacts of a service component);
- 3) 'normative service models' – organizationally focused marketing models (like incentive schemes and trade-offs between satisfaction and productivity), and operations models (like queuing).

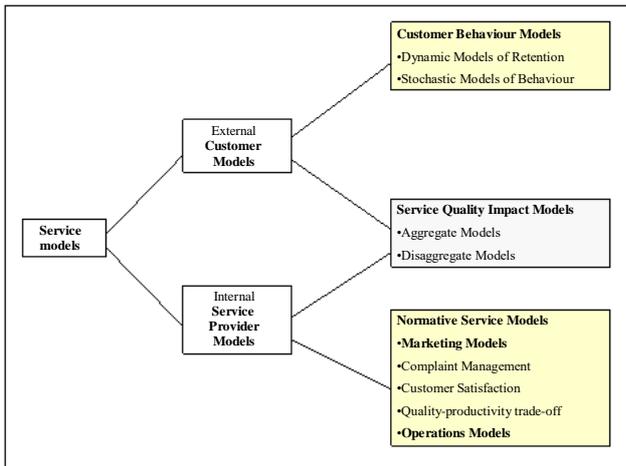


Figure 1. Mathematical Models of Service, from Rust & Metters (1996)

In 1999, Cook Goh & Chung[20] developed the 'integrated schematic representation of services' matrix. Figure 2 displays this integrated services schematic.

They recognized that services could be split into marketing (product) or operations (process) orientations. They believed that in delivering a final 'customized' solution their remained a need to integrate and interact with both orientations. They suggested research in the 'interaction and integration' area may articulate strategies and tactics for improving services.

Roth & Menor[21] delivered a further addition to the services topologies. Their 'service strategy triad' (displayed in Figure 3) separated the 'what', the 'how', and the 'who' of service encounters. It offered a new

perspective to advance an understanding of services operations management. The 'who' was the right customers (and not just a customer segment). These targeted customers could be defined by techniques like Forrester's 'technographics', 'psychographics, and psychographic profiling groups of customers. The interpretation of such target markets provided a means to enhance both service and performance standards, and to allow the business to competitively align its chosen degree of customer targeting with its offered service products and delivery systems.

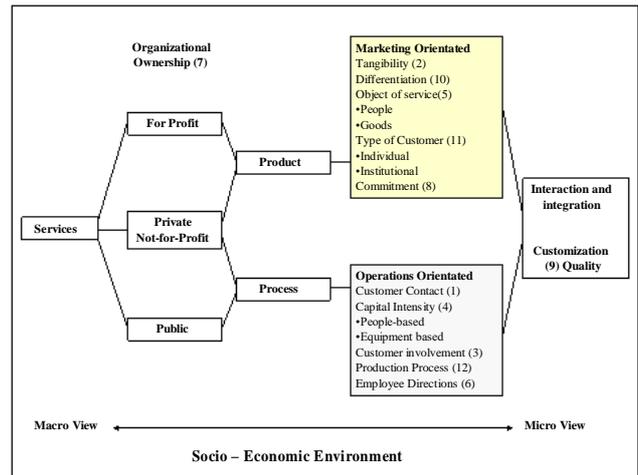


Figure 2. Integrated Schematic Representation of Services, from Cook, Goh & Chung (1999)

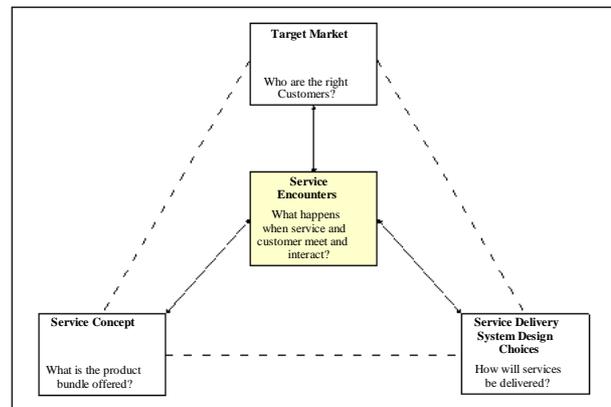


Figure 3. The Service Strategy Triad, from Roth & Menor (2003)

Roth & Menor operationally defined their 'service strategy triad' into 5 elements:

- 1) the supporting facilities (physical and structural resources);
- 2) the facilitating goods (materials and supplies that are consumed);
- 3) the facilitating information (supporting the explicit services);
- 4) the explicit services (customer experiential and sensual benefits);
- 5) the implicit services (psychological benefits).

They realized the total service concept by the customer, may differ from the service offered by the service provider. To overcome this, a feedback loop (execution, assessment of gaps, renewal) was proposed.

Roth & Menor’s ‘service delivery systems architecture’ model, displayed in Figure 4, allows a framework to investigate three interrelated and dynamic components of service delivery systems:

- 1) the strategic service design (portrayed as structural, infrastructural and integration, and based on choices between time-phased content portfolios of major supply
- 2) the service delivery execution system (exemplified by programs, policies and behavioural aspects deliver complimentary areas of customer focus], possibly using balanced scorecard approaches;
- 3) the customer perceived value of the total service concept (intangibles and other effectiveness aspects of the service).

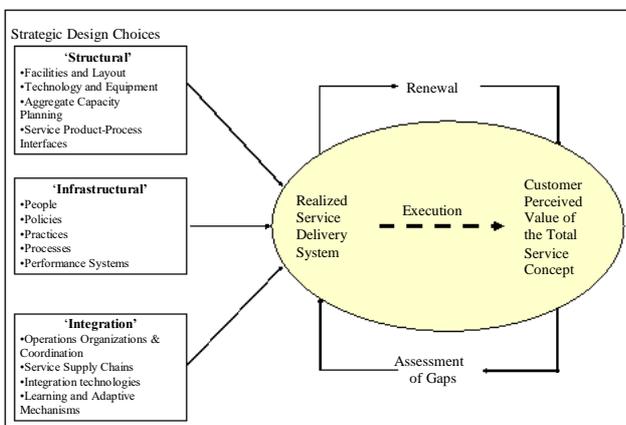


Figure 4. The Service Delivery Systems Architecture, from Roth & Menor (2003)

These features, delivered upstream by the external integration of the service supply chain, combined with the linked internal integration of the operational functional areas, and the adaptive mechanisms available to the intellectual capabilities provide new avenues to perceived customer value. Here the artificial learning and intelligence capabilities of the system, the fuzzy logic approaches to approximate answers, and the rapid absorption of skilled human capital talents may deliver targeted boundaries which offer enhanced potential to the service encounter. The assessment execution and renewal factors encountered define the potential effectiveness of the realized service encounter. These choices are the basis for forming competitive capabilities, in such strategic arenas as the internet, the manufacturing arena, internet marketing, and the services area.

The degree of service competitive capabilities such as consistent quality, supply chain management, convenience, accessibility to channels, customization, one-on-one customerization, operational efficiency and low costs are used by customers to make choices among competitors.

The above models, when considered with other factors like customer demand-driven supply chains, the effect of the internet, business solutions, virtual service organizations and technology options, along with their interrelated, and interconnected links, can be drawn into a new topology model termed the ‘service value-chain encounter model’.

4. SERVICE VALUE ENCOUNTER MODEL

The ‘service value-chain encounter model’ offers the framework for a new research model. This model is displayed in Figure 5. The operational, services and customer strategies of the business are drawn together as interconnected data sharing models delivering unique customer services encounters – ones aiming to exceed customer expectations! This business system learns from its customer encounters, and improves its services database offerings ready for additional, or more specific, customer encounters.

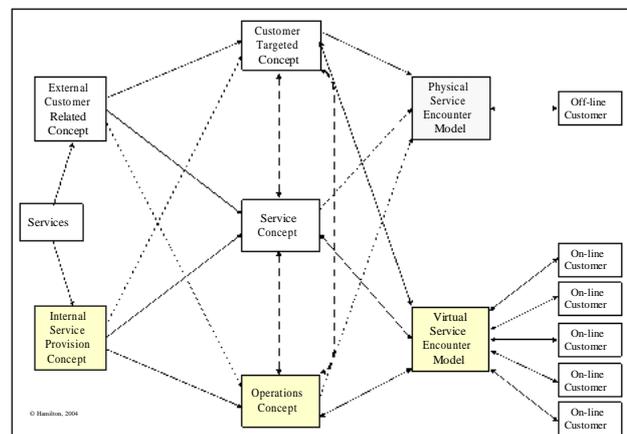


Figure 5. The Physical & Virtual Service Value Encounter Model © Hamilton (2004)

This model meshes with:

- 1) earlier works above;
- 2) other works[22],[23],[24];
- 3) commercial services business solutions providers [25],[26];
- 4) software developers including Microsoft & IBM.

The ‘operations concept model’ component delivers the business’s latest service value-chain networked information systems, data storage and retrieval systems. It incorporates latest web metrics[27], fuzzy logic[28] and artificial intelligence[29] tools, to:

- 1) interrogate its internal and external databases;
- 2) sort and interpret available information;
- 3) deliver customized (or personalized) solutions capable of targeting perceived physical or virtual customer expectations.

This ‘renewal’ and learning networked system is very different to the normal ‘rigid core component’

experienced in normal website service offerings[30]. It also incorporate issues related to failures and recovery [31]. In short, the 'operations concept model' component delivers the 'how' to the service value encounter.

The 'services concept model' component is integrated with both the operations concept model and the customer targeted model and delivers multi-dimensional information from the business and its value adding partners, in a quality manner, such that the greater the depth of information sought, the greater the quality, and the more realized[32] is the customer's experience. In short, the 'services concept model' component delivers the 'what' to the service value encounter

The 'customer targeted model' component establishes a relationship to the service being offered in terms of its efficiency, relevance, scope and performance[33]. This requires the cognition to recognize, and then target, the specific customer group. In some cases it allows for degrees of customization (or even one-on-one 'customerization'), of the customer service product. The understanding of the target market(s) remains an important consideration as it enables a viewpoint for new service development [34]. In short, the 'customer targeted model' component delivers the 'what' appropriately targeted to the 'whom'.

These three component models and their downstream areas house much of the business's intellectual property. They remain integrally linked downstream via a modern 'service value-chain'. Downstream business supply chain partners, logistics support, peripheral partners and other external data sources are interconnected and interrogated across the business integrated IT networks. Here internal and external data, logistics, and the like are pooled, shared, cross-model, and then applied to provide new business systems learning, and new, improved, upstream customer solutions. The customer and the business network come together at the service encounter 'touch-point' Thus, the business aims to deliver the 'best', customer demand chain driven, customer value chain response, possible from the available value adding set of databases at its disposal.

The 'service encounter' has two components –

- 1) a 'physical' (tangible) encounter between the customer and a business contact person or persons;
- 2) a 'virtual' (intangible) encounter with and electronic based structure, which is often visually connected via its internal or external business website. In both cases information flows from customer to business and business responds sourcing relevant, allowable (non-sensitive), correct information.

Business then delivers customer requested, value chain sourced information (in a timely flow), across the service encounter interface and through to the customer. Chinese University of Hong Kong and National Sun Yat-Sen University research [35] indicates customers in the US

buy using multiple channels - including stores, catalogues, and on-line activities. The service value-chain operates across both physical and information pathways and networks. It is a key part of the operations management equation. Such viewpoints support the physical and virtual service value encounter model. Sterne [36], a world authority on web metrics, supports the notion that businesses offering both the physical and virtual encounter options tend, in the current market to be more successful, but that both models may also operate independently.

Businesses today are increasingly developing extensively networked on-line offerings, combined with high levels of interconnectivity between partners, alliances, and associated value adding organizations. In addition, they are moving their supply chains into high-tech, networked, intelligent solutions – termed service value-chains [37]. The 'service value-chain' may be defined as the flexible, dynamic, delivery of a service, or product, by a business's coordinated value chains (supply chains and demand chains working in harmony), such that a value-adding, specific, service solution is effectively, and efficiently, delivered to the individual customer in a physical or virtual manner.

In 2003, Australia's 'business to business' and 'business to consumer' e-commerce was valued at \$11.3B [38], and it was growing rapidly. It ranked 5th in the world regarding its potential to use the internet economy! At this time, thirty five percent of Australian businesses purchased on-line and eighty nine percent of Australian businesses were on-line. Australian businesses (with more than ten employees) recorded near ubiquitous internet adoption [38]. The internet has transformed many of Australia's key business and agency functions including services delivery, customer relationship management, organizational administration, supply chain management and knowledge (or data) management. In January 2004 over forty six million web servers worldwide were globally connected to the internet, with

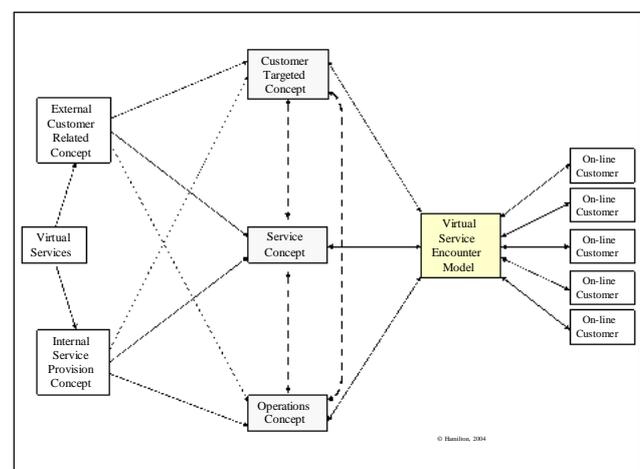


Figure 6. The Virtual Service Encounter Model © Hamilton (2004)

ninety six percent connecting with the browser Internet Explorer. Thus, for many countries like Australia the opportunity remains to deliver high-value service offerings to virtual business customers.

The business website presents one form of the ‘customer service value-chain encounter model’ – the ‘virtual service encounter model’, and it is displayed in Figure 6. This e-service encounter environment presents several potential virtual customer services related weaknesses and several points of research including:

- 1) amplification effects;
- 2) e-services;
- 3) the web interface;
- 4) value chain modelling;
- 5) customer targeting;
- 6) information communication technologies;
- 7) bottleneck effects;
- 8) business strategies.

These factors are elaborated below.

4.1 Amplification Effects:

The bullwhip effect is defined as the ‘phenomenon where orders to the suppliers tend to have larger variances than sales to the buyer (demand distortion) and the distortion moves upstream in a amplified form (variance amplification)[39]. Lee, et al.[40] consider customer ordering as a lumpy occurrence to which the supplier responds. Schmenner[41] theorized that fast, even-flow could explain productivity gains in manufacturing settings, and that even flows were achievable when the variability in supply chain was reduced. He also noted that tight quality control reduced negative amplification effects. Finch[42] states; ‘a key to eliminating the bullwhip effect and a key to any supply chain management effort is an increase in ‘information’ supplied by business to their suppliers’. Thus the upstream amplification (or bullwhip) effects may be reduced where a website delivers more efficient, more direct, targeted, more uniform, information access pathways between the service value-chain and the customer. Improved information flows - delivering manageable amounts of filtered most relevant information to the customer may deliver one solution.

4.2 E-Service Effects:

Many market forces influence the development of the service value-chains. There is an ever-present economic imperative to reduce IT costs, whilst increasing both the business value, and impact of this IT suite. Many businesses cannot afford high degrees of IT customization, and indeed this may not always be necessary – consider a mass user situation like on-line airline bookings. Hence size and capital remain limiting factors. Timelines to move to on-line service value-chain solutions also vary.

A true service value-chain integrates all aspects of its business’s service supply chain – internal and external in an intelligent, coordinate manner. It then interrogates the relevant data and delivers business-specific intelligence that matches the demands of the customer, again reducing inefficiencies. This area remains one with great scope for further research and development. Various third party logistics solutions have been developed to integrate these areas with those of other businesses, delivering new strategies, solutions and competitive advantage.

4.3 Website Effects:

Incorporating software programs that query the available database information apparently ‘intelligent’ solutions to customer requests may be offered.

The intelligent website acts similarly to an intelligent, inquisitive, reasoned, language sensitive search engine, capable of taking in customer requests by voice, email, image, ‘search-for ...’, and the like. Artificial intelligence, knowledge management, and fuzzy logic principles are then applied to determine efficient, appropriate business-specific solutions.

The resulting ‘intelligent’ website processes may reduce the need to revisit and reinterrogate databases, and may reduce the non productive, time consuming, information seeking workload requirements on the service supply chains. Significant, tangible and intangible cost reductions (less non productive website activities) may be achievable. Less demands, per initiative, per customer, may reduce the information transmission strain across the global communications networks, and may reduce negative customer sentiment.

Such solutions may require the flat website encounter to move to a new three dimensional approaches[43].

4.4 Value Chain Modelling Effects:

Value chain modelling[44] has shown how changes in speed, responsiveness, and variability affect operational performance, and may enable the business to perform solutions scenario like financial impact assessment, cost-benefit analysis, and sensitivity analysis.

Vermijmeren[45] suggests flexible, intelligent supply chain ‘engines’ may drive these dynamic supply chains, delivering value, in an efficient manner. Various third / fourth party logistics additions offer additional new strategic solutions, scope for competitive advantage, and scope for business development.

The incorporation of peripheral added value options may improve these offerings.

4.5 Customer Targeting Effects:

When a customer encounters any aspect of a business a 'moment of truth' arises, and positive or negative impressions can be generated[46]. In highly customer responsive business systems, customer contact time may be lessened and sales opportunities may be enhanced. The customer may be an internal customer (working for the business, a participant in the upstream service value-chain, or an internal services participant in an area such as data processing, engineering, maintenance, accounting, after sales service[47], or an external customer (a consumer or one who interacts with and adds value to the business service value-chain. To service the virtual customer complete, responsive, flexible, adaptive, service value-chains often offer most desirable options. Finch argues that delivering quality services involves assurance and empathy. However to deliver service quality across a website requires a broadening of the quality dimensions to encapsulate product and service and product dimensions. Thus the customer exhibits a multidimensional impact on the business website. Hence, the business must maximize its virtual 'touch-point' appeal, and must develop its virtual management tools (and strategic metrics) set.

4.6 Information Communication Technology Effects:

In 2001, the five largest software providers: HP, IBM, Microsoft, Oracle, and Sun, along with a few new entrants, began promoting new standards, new web services platforms, and new activities environments. Since 1998 internet protocol version 6 (IPv6) software has been encapsulated into operating systems platforms[48]. If adopted, IPv6 will allow marketers to segment a business's website customers using postcodes, geographical location, and phone numbers. Mobile devices, watches, and clothing, are now capable of housing customized information solutions for business. Third party software operating on common platforms like Microsoft's '.Net', or IBM's WebSphere platforms may further enhance the virtual environment delivering savings for business measured as per initiative - lower 'human', and 'capital resource requirements'. In addition, new ways to interpret, interrogate and deliver customer requirements are unfolding, and highly intelligent, responsive websites are emerging. New business strategies, and the nuances of customer wants and needs are developing and will be incorporated into solutions. Working relationships - like 'e-customer relationship management', trust, loyalty, satisfaction, addressing the dynamics of the industry structure, and cultural fit will become just as important to the customer as the provider's portfolio. Currently latest computer application tools deliver low level solutions (not high levels of customization), and are best utilized for standard product type applications. Hence the development (and implementation) of highly customerized, and fully operational, 'service value-chains' is not yet a reality.

4.7 Bottleneck Effects:

Bottlenecks occur when a limiting resource affects the output level of the entire system. The business-customer website encounter as shown in Figure 6, is one such bottleneck. Here, multiple customers search multiple supply chain data sources, for their individual needs, and business inefficiencies arise. Finch suggests bottlenecks may be considered as business constraints.

In the services industry, information is the key ingredient that moves. Some information may be physical in nature - like paperwork, whereas in the manufactured product situation, both information, and product move. In both cases information bottlenecks occur.

Efficient design of the website (with the use of appropriate technologies), may reduce customer cycle-time (customer website access time to source, retrieve and absorb desired information) [49],[50], reduce bottlenecks, and possibly improve website effectiveness. Thus 'touch-point' information trade-rates between the customer and the business service value-chain remain dependent variable areas that may be improved.

4.8 Business Strategies Effects:

Businesses faced with tough competition are devoting greater resources to support their e-business initiatives [51]. Using tools defined by IBM and others these businesses can prioritize their financial and operational performances, and closely define their e-business and management strategies in multiple-customer environments. These strategies involve the development of semi-intelligent websites, and target delivering in four key areas related to the 'physical and virtual service encounter model' - these are the: 1) technical (operations model) factors - including communication channels; software and hardware; artificial intelligence; fuzzy logic methodologies; natural language interpretations; web metrics; website flows and information integration; presentation modes (3D screens); and telepresence. 2) business (service operations model) factors - including externals - supply chain partners; peripheral partners; logistics; and internals - business, management; marketing; operations; and strategies. 3) customers (customer targeting model) factors - grouped initially in technographic segments, then cyber-segmented further to eventually allow individual targeting. 4) revenue generator factors - including - sales, fees, charges, advertising, partnerships, and franchises.

5. BUSINESS RAMIFICATIONS

An understanding of the 'virtual service encounter model', and its 'effectors' provides the business strategist with a new disruptive business model. This model facilitates the development of inter-service-model concept measurements, and relevant integration processes.

Each business has a unique mix of effectors and further research in this area is currently being pursued by the author. Investigation is continuing on 5 fronts – customer targeted constructs, on-line customer constructs, serviced concept constructs, operations constructs and financial constructs. The aim of this research is to deliver a new customer touch-point model, one that delivers greater satisfaction to the virtual customer and does so in a more efficient manner, by better developing and utilizing the service value-chains. This will ultimately deliver a highly agile, flexible service value network, and do so at reduced financial commitment per virtual customer.

6. CONCLUSION

The ‘physical and virtual service encounter model’ provides a service value-chain view of areas where a business may attempt to close the gap that exists between ‘business issues’ in the client side and ‘technology capabilities’ in the provider side. The virtual customer ‘touch-point’ is complex, and many areas of research, development and idea creation may be applied to enhance the business’s website competitiveness [52].

Hence the realm of business, its virtual service value-chains, and its virtual customer remains an exciting area of investigation. This paper delivers many related opportunities for research and business development and scholars are encouraged to join in developing this emerging field of research and disruptive technologies.

7. REFERENCES

- [1] Hughes, J. Mitchell, P. & Ramson, W. (eds), 1993, *Australian Concise Oxford Dictionary*, Oxford University Press, New York, NY.
- [2] Kathawala, Y. & Abdou, K. 2003, ‘Supply chain evaluation in the services industry: A framework development compared to manufacturing,’ *Managerial Accounting Journal*, vol. 18, no. 1/2, pp. 140-149.
- [3] US Bureau of Labor Statistics, 2002; 2004.
- [4] Kamarkar, U. 2002, ‘Integrative Research in Marketing and Operations Management,’ *Journal of Marketing Research*, vol. 33, no. 2, pp. 125-133.
- [5] Chase, R., Aquilano, N. & Jacobs, R. 1996, *Operations Management for Competitive Advantage*, McGraw-Hill/Irwin, Boston, MA.
- [6] Cook, D., Goh, C. & Chung, C. 1999, ‘Service topologies: A state of the art survey,’ *Production and Operations Management*, vol. 8, no. 3, pp. 318-338.
- [7] Nie, W. & Kellog. D. 1999, ‘How Professors of Operations Management View Service Operations,’ *Production and Operations Management*, vol. 8, no. 3, pp. 339-355.
- [8] Judd R. 1964, ‘The case for redefining services,’ *Journal of Marketing*, vol. 28, no. 1, pp. 58-59.
- [9] Rathmell, J. 1974, *Marketing in the Service Sector*, Winthrop Publishing, Cambridge, MA.
- [10] Quinn, J., Baruch, J. & Paquette, P. 1987, ‘Technology in Services,’ *Scientific American*, vol. 257, no. 6, pp. 50-58.
- [11] Quinn, J., Baruch, J. & Paquette, P. 1987, ‘Technology in Services,’ *Scientific American*, vol. 257, no. 6, pp. 50-58.
- [12] Zeithaml, V., Berry, L. & Parasuraman, A. 1988, ‘Communication and Control Processes in the Delivery of service Quality,’ *Journal of Marketing*, vol. 52, no. 2, pp. 35-48.
- [13] Czinkota, M., Ronkainen, I. & Moffett, M. 2005, *International Business*, (7th ed.), Thompson/South-Western, Mason, Ohio.
- [14] Berry, L. & Parasuraman, A. 1991, *Marketing Services: Competing through Quality*, The Free Press, New York, NY.
- [15] Levitt, T. 1972, ‘Production line approach to service,’ *Harvard Business Review*, vol. 50, no. 5, pp. 41-52.
- [16] Rust, R. & Metters, R. 1996, ‘Mathematical Models of Service,’ *European Journal of Operations Research*, vol. 91, no. 3, pp. 427-439.
- [17] Verma, R. & Boyer, K. 2000, ‘Service Classification and Management Challenges,’ *Journal of Business Strategies*, vol. 17, no. 1, pp. 5-24.
- [18] Akkermans, H. & Vos, B. 2003, ‘Amplification In Service Supply Chains: An Exploratory Case Study From The Telecom Industry,’ *Production and Operations Management*, vol. 12 no. 2, pp. 204 -223.
- [19] Chen , I. & Paulraj, A. 2004, ‘Towards a theory of supply chain management: the constructs and measurements,’ *Journal of Operations Management*, vol. 22, pp. 119-150.
- [20] Cook, D., Goh, C. & Chung, C. 1999, ‘Service topologies: A state of the art survey,’ *Production and Operations Management*, vol. 8, no. 3, pp. 318-338.
- [21] Roth, A. & Menor, L. 2003, ‘Insights into Service Operations Management: A Research Agenda,’ *Journal of Operations Management*, vol. 12, no. 2, pp. 145-164.
- [22] Hoffman, D. & Novak, T. 2000a, ‘Advertising Pricing Models for the World Wide Web,’ in D., Hurley, B., Kahin, & H., Varian, (eds.) *Internet Publishing and Beyond: the Economics of Digital Information and Intellectual Property*, MIT Press, Cambridge, UK.
- [23] Chen, Q. & Wells, W. 1999, ‘Attitude toward the site,’ *Journal of Advertising Research*, vol. 39, no. 5, pp. 27-38.
- [24] Biocca, F., Li, H. & Daugherty, T. 2002, ‘The Role of Virtual Experience in Consumer Learning,’ *Journal of Consumer Psychology*, vol. , no. , pp. 1-39.
- [25] Gartner (www.Gartner.com) Gartner Dataquest (April 2002).
- [26] Comergent.com. Retrieved: April 10, 2004, from <http://www.comergent.com/products/valuepropositions.cfm>.
- [27] Sterne, J. 2002, *Web Metrics*, John Wiley & Sons, New York, NY.

- [28] Louvieris, P. & Driver, J. 2001, 'New frontiers in cyberssegmentation: Marketing success in cyberspace depends on IP address,' *Qualitative Market Research: An International Journal*, vol. 4, no. 3, pp. 169-181.
- [29] Kemp, R., Schot, J. & Hoogma, R. 1998, 'Regime shifts to sustainability through processes of niche formation: the approach of Strategic Niche Management,' *Technology Analysis and Strategic Management*, vol. 10, no. 2, pp. 175-195.
- [30] Leonard-Barton, D. 1995, *Wellsprings of Knowledge*, Harvard Business School Press, Boston, MA.
- [31] Miller, J., Craighead, C. & Karwan, K. 2000, 'Service Recovery: a framework and Empirical Investigation,' *Journal of Operations Management*, vol. 18, pp. 387-400.
- [32] Mintzberg, H. 1978, 'Patterns in Strategy Formation,' *Management Science*, vol. 24, no. 9, pp. 934-948.
- [33] Srinivasan, J., Jagannathan, S. & Kalman, J. 2002, *Internet Commerce Metrics*, Prentice Hall, Upper Saddle River, NJ.
- [34] Heskett, J., Sasser, W. & Hart, C. 1990, *Service Breakthroughs – Changing the Rules of the Game*, The Free Press, New York, NY.
- [35] Liang, T. 2003, 'The Second Wave of e-Commerce: Successful Transformation and Business Models,' Keynote Address, 3rd *International Conference on Electronic Business*, Singapore.
- [36] Sterne, J. 2002, *Web Metrics*, John Wiley & Sons, New York, NY.
- [37] Barlow-Hills, S. & Sarin, S., 2003, 'From Market Driven to Market Driving: An Alternative Paradigm for Marketing In High Technology Industries,' *Journal of Marketing Theory and Practice*, vol. 11, no. 3, pp. 13-24.
- [38] Di Gregorio, J. & De Montis, F. 2003, 'NOIE Information Economy Index 2003,' *National Office for the Information Economy*, pp. 1-50. Retrieved: March 12, 2004, from http://www2.dcita.gov.au/_data/assets/file/12294/NOIE_Index.pdf.
- [39] Forrester, J. 1961, *Industrial Dynamics*, MIT Press, Cambridge, MA. & Akkermans, H. & Vos, B. 2003, 'Amplification In Service Supply Chains: An Exploratory Case Study From The Telecom Industry,' *Production and Operations Management*, vol. 12 no. 2, pp. 204 -223.
- [40] Lee H., Padmanabhan & Whang, S. 1997, 'The Bullwhip Effect in supply Chains,' *Sloan Management Review*, Spring, pp. 93-102.
- [41] Schmenner, R. 1995, 'How Can Services Businesses Survive and Prosper?,' *Sloan Management Review*, vol. 27, no. 3, pp. 21-32.
- [42] Finch, B. 2003, *OperationsNow.com*, McGraw-Hill/Irwin, New York, NY.
- [43] Microsoft at Microsoft.com.
- [44] Bagchi, S., Gallego, G., Grey, W., Katircioglu, K., Seybold, D., Shi, D. & Stefanis, S. 2003, 'An analytical approach for quantifying the value of e-business initiatives,' *IBM Systems Journal*, vol. 42, no. 3, pp. 484-497.
- [45] Vermijmeren, M. 2003, 'Software architecture in supply chain management,' *Computers in Industry*, vol. 53, pp. 165-178.
- [46] Albrecht, K. & Zemke, R. 1985, *Service America! Doing Business in the New Economy*, Dow Jones-Irwin, Homewood, H.
- [47] Davis, M, Aquilano, N. & Chase, R. 2003, *Fundamentals of Operations Management*, (4th ed.), McGraw-Hill/Irwin, New York, NY.
- [48] Comer, D. 1995, *Internetworking with TCP/IP volume 1: Principles, protocols, and architecture*, Prentice-Hall, Englewood Cliffs, NJ.
- [49] Malcinski, A., Dominick, S. & Hartrick, T. 2001, 'Getting to know your audience is the key to online success – making the numbers work for you,' *IBM white paper*, pp. 1-11. Retrieved: May 1, 2004, from www-106.ibm.com/developerworks/web/library/wa-mwtl/.
- [50] Cutler, M. & Sterne, J. 2002, *E-Metrics*, NetGenesis, pp. 1-62.
- [51] Bowman, R. 2001, 'Fast deployment - quick return help sell IT in a slow economy,' *Global logistics and Supply Chain Strategies*, Oct.
- [52] Porter, M. 2001, 'Strategy and the Internet,' *Harvard Business Review*, Vol. 79, No. 3, pp. 63-78.

This paper is further developed in the JSSSE 2005

The Web Services on Mobile Shopping Automation

Chengchih Yang¹, Chengmin Lin², Shihung Chen³

¹Department of Information Engineering

²Department of Computer and Communication Engineering

³Department of Multimedia Design

Nan-Kai Institute of Technology, Nan-Tou, Taiwan, China

{czyang, lcm, and sostc}@nkc.edu.tw

ABSTRACT

This paper presents the Web services applying on a shopping automation across the Internet from portable mobile devices accessed by users. First of all, a shopping automation is considered as hardware system such as embedded device products. They are binding with web in some protocol. Another hand, the portable mobile devices also connected with web in WAP protocol. We model them in Web Services Description Language (WSDL) to represent such environment system. An example is described in this study that includes a General Packet Radio Service (GPRS) platform is used as the function of home gateways. In addition, an XML-based protocol is defined as the interface between web server and shopping automations.

Keywords: shopping automation, WAP, WSDL, GPRS, XML-based protocol

1. INTRODUCTION

Web services are more important with increasing the number of portable mobile devices, such as mobile phones, Personal Digital Assistant (PDAs). A typical Web services architecture consists of three entities, including service providers, service requesters, and service brokers [9]. Service providers create Web services for service requesters. These Web services are registered in service brokers and the brokers handle the requesting from service requesters and find published Web services of service providers. Hence, efficient access to Web services became an important issue. Ouzzani and Bouguettaya provide a query optimization model based on aggregation the quality of Web service parameters of different Web services [8]. In addition, Web services consist of three important protocols. First of all, Simple Object Access Protocol (SOAP) provides a simple and lightweight protocol for communicating between Web services and client applications. Next, Web Services Description Language (WSDL) is an Extensible Markup Language (XML) based language to be used for describing Web services. Furthermore, Thomas et. al use Petri Nets to model the WSDL [13]. Finally, the Universal Description Discovery and Integration (UDDI) is used to let Web services register their services for services requesters. In order to test and evaluate Web services, Tsai et. al provide an XML-based Framework that consists of test master and test engine [12]. The former allows testers to specify test scenarios and cases converted from WSDL specifications. After the Embedded Internet System (EIS) was proposed in 1998 [7], academic researchers in this field are increasing in number. For example, Szymanski [11], designed web that applied the EIS ideas to enable control webs to perform online monitoring. The Open Services Gateway Initiative (OSGi) [5], whose mission is to generate open

specifications for delivering of various services over wide area networks to local networks and devices was founded in March 1999. The three core aspects of the OSGi mission are servers, various networks, and devices. In 2000, Topalis et al. implemented a home gateway to support automated home networks and home Internet connectivity [14]. In 2001, Saif et al. proposed a self-configuring software architecture for home area networks, which provides an extensible markup language (XML) based registry and a hypertext transmission protocol (HTTP) based event service [10]. In 2001, Filman applied an embedded computer to design a home gateway, because embedded computers had progressed from being simple machines to being "thinking" and "sensing" machines [3]. Furthermore, Chen et. al in 2002 provide an XML-based graphical monitoring protocol for distributed systems [2]. In 2003, Kapsalis et. Al [6], used the technologies Object Link for Process Control (OPC) enable Internet access, and proposed a gateway to integrate various services. Although a large number of studies have been made on both Web service and home automations or shopping automations, little is know about integration of both. The first aim of this study is to develop a web-based monitoring system for providing Web services. The remainder of this article is organized as follows. Section 2 presents the proposed system architecture. Section 3 introduces this Web representation in WSDL and a simple protocol is defined as the interface between automation server and shopping automations. A mobile shopping automation example is illustrated in Section 4. Finally, Section 5 draws conclusions.

2. SYSTEM ARCHITECTURE

A three-tier application has front-end clients (e.g., browsers), middle-tier application servers (e.g., web servers), and back-end databases. It is one where the

logical decomposition of the application is reflected both at the software and hardware level. Similarly, a mobile shopping automation system, including portable mobile devices, an application server, a database, an automation server, and a shopping automation network, is now constructed and show as Fig.1. The mobile device, the application server, and the database server can be regarded as a data process path, another way, the database server, the automation server and the shopping

automations can be regarded as being a signal process path. They have a common database for data process path and signal process path. Although such an integrated model does not yet have a common standard for mobile monitoring, the definitions of among layers are useful to clarify rules and developments of the interface protocol. The server described above is used to perform the functions of an application server and automation server.

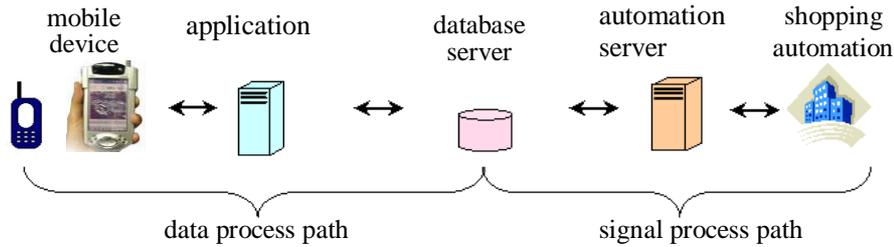


Fig. 1. A mobile shopping automation architecture

3. WEB REPRESENTATION

Web representation is the main goal of this work. We'll analysis those operations and protocol parameters in mobile shopping automation. For Applying WSDL, a WSDL flow model is used to model our system to descript our Web services. Finally, an automation network protocol in our system is described herein.

3.1 Operations and protocol parameters

Three operations are provided from application server to mobile user, including inquire, control, and push. A user can call on a web site to retrieve the sensor status T_s in the database. This procedure is called the inquire operation and is denoted as O^q . For the database, O^q is a read-only operation and so is a concurrent process. Next, a user can change the sensor status, and use a control operation to finish it. This control operation is denoted as O^c . It is not a concurrent process. This operation must cooperate with a mutual exclusion table to ensure consistency among data. Although database operations provide a transaction mechanism, an inconsistent state may occur. For instance, two users, A and B,

simultaneously change the status of a single sensor. The successive operations are $O_a^q O_b^q O_a^c O_b^c$ and an inconsistency arises. An automation server is built to prevent the above inconsistency. However, the server can process concurrent problems. The data process path can simply prevent inconsistency because the server provides a mutual exclusion mechanism. The last operation associated with the data process path is a push operation. For a mobile system, two models can be applied, *pull* and *push*. For push operation, the application server actively sends information to a user's mobile device. SMS or MMS technology is employed to implement the push mode. Such an operation is denoted as O . Two operations are provided from automation server including *in* and *out*. The automation server extracts a control command from table T^c of the database and then sends the command to the sensor. The procedure to be used is denoted as O Moreover, the server periodically reads the sensor status and then stores the status into the status table T and a message table T^m . This procedure is called *in* and is denoted as O^i . Then we draw them into the system block diagrams and show in Fig.2.

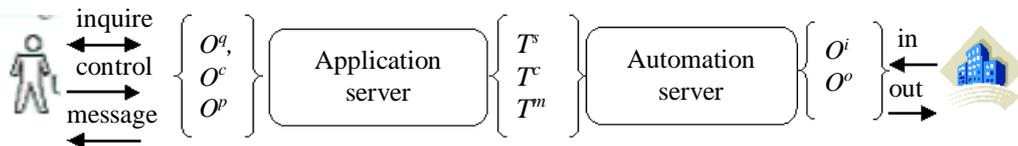


Fig.2. operations on mobile environment

Furthermore, we define a mobile control web operation set O^{mcw} , includes five primary functions. $O^{mcw} = \{O^q, O^c, O^p, O^o, O^i\}$. For the inquire operation O^i used to represent an object embedded some parameters includes name, value, and protocol fields. The name filed is the name of the object. It consists of sixty-four characters and is unique to represent a primary key for

this object. The value field is also string type because it can represent a value of float and integer. In addition, using string also represent a binary status into {on, off}. Protocol field parameter P^s is used to carry application server communication with client. For operations O, O^i , the parameter of protocol field P^s is used to carry automation server communication with shopping

automations. We also define a protocol set P^{mcw} of a mobile control web includes two protocol parameters P^h , P^g .

3.2 WSDL flow model

WSDL is an XML-based document includes specific elements: types, message, port-type, binding and Service. It is used to describe what a web service can do, where it resides, and how to invoke it. In this study, we map the mobile shopping automation environment to WSDL as shown in Fig. 3. In this figure, there are two ports, port-c and port-s. The former uses WAP protocols to bind between a mobile client and an application and the latter is used to bind a sensor network and an automation server using sensor protocols. The protocol

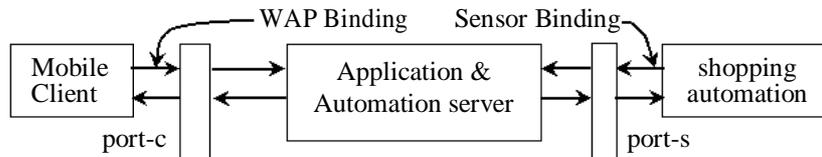


Fig.3. A mobile home automation for WSDL flow

Fig.4. The DTD of SN protocol

3.3 Sensor Network Protocols

In this section, we describe sensor network (SN) protocols in our system. The Document Type Definition (DTD) of the SN protocol is shown in Fig.4. There are two types of command elements: Digital Data In (DI), and Digital Data Out (DO) in this figure. Each command element has a device attribute. If a command is sent from a GPRS control platform to an appliance server, the device attribute is used to specify as the identification (ID) of the target device. If the command is sent from the appliance server to the GPRS control platform, the device attribute is used to store the ID of the source device.

```
<?xml version="1.0" encoding="UTF-8"?>
<!ATTLIST DI
  device ID #REQUIRED
  value (On | Off | Query) #REQUIRED>
<!ELEMENT DO EMPTY>
<!ATTLIST DO
  device ID #REQUIRED
  value (On | Off) #REQUIRED>
<!ELEMENT DI EMPTY>
<!--root element-->
<!ELEMENT SN (DI | DO)+>
```

4. EXAMPLES

The information server is constructed using server operation systems such as Windos2000 Server or Windos2003 Server, and application software for web pages and databases. A remote appliance monitor is composed. It has a control platform MA8-6 that

is defined in next subsection. In spite of port-c or port-s are used in GPRS wireless channels. When a request is send to the application server. Herein, the application is to play as a service broker. Hence, the application can find a suitable automation server from its Database registered by automation servers. When a application server is selected, it sends a query command to a specify sensor device in a sensor network through a GPRS channel. In our system, a sink node introduced by Akyildiz *et. al* [1] can receive the command and process this command to read the status of the specify sensor device. Next, the information of the specify sensor device backs to the automation sever by the sink node. Finally, the automation server provides this service, a required sensor status, to the mobile client that proposes this request.

contains multi-type I/O interfaces, DIO, RS232 or RS485, linked to the web's appliance server via GPRS protocols. This system can be used as a set of sensors to detect earthquakes, stealers, or fires. Appliance server access several numerous detector data and save to the public databases for information server sharing. The main effect of the application server is just to provide a html or asp.net webpage to supply an ideal interface between the web's database and the user of a cellular phone. The user can enquire about, and control the appliances from his or her cellular phone through the WAP or HTTP protocol. Figure 5 shows a sensor network that consists of anti-fire, anti-earthquake and anti-stealing circuit-broad, and the GPRS control platform, MA8-6. Figure 6 shows the states of the appliances from database using an IE browser.

5. CONCLUSIONS

This study addresses an integration of web services apply on mobile shopping automations. We hypothesize that it can be treated as a both software and hardware architecture in order to model the relation between Web services and mobile shopping automation. The servers provide mobile Web services and automations controls. They are important roles in our system. The application server is a service broker, the automation server is a service provider, and a mobile client, mobile phone or PDA, is service request. The architecture of Web services for mobile shopping automation is represented. In addition, the GPRS mechanism and protocol of SMS can be used to achieve real-time monitoring functions for remote monitoring.

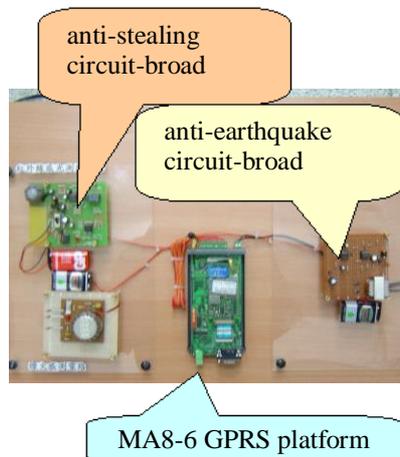


Fig. 5. GPRS platform, MA8-6

ACKNOWLEDGEMENT

The authors would like to thank the National Science Council of Taiwan, China for financially supporting this research under Contract NSC-91-2213-E-252-008.

REFERENCES

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless Sensor Networks: a Survey," *Computer Networks*, no. 38, pp. 393-422, 2002.
- [2] S. S. Chen, C. M. Lin and S. S. Wu, "An XML-based Graphical Monitoring Protocol for a Distributed Graphical Monitoring System," in the Proceedings of 2002 International Conference on Chinese Language Computing, 2002.
- [3] R. E. Filman, "Embedded Internet Systems Come Home," *IEEE Internet Computing*, vol. 5, no. 1, pp. 52-53, Jan./Feb. 2001.
- [4] S. Frolund and R. Guerraoui, "e-Transactions: end-to-end reliability for three-tier architecture," *IEEE Transactions on Software Engineering*, vol. 28, no. 4, pp. 378-395, 2002.
- [5] L. Gong, "A Software Architecture for Open Service Gateways," *IEEE Internet Computing*, vol. 5, no. 1, pp. 64-70, Jan./ Feb. 2001.
- [6] V. Kapsalis, K. Charatsis, M. Georgoudakis, and G. Papadopoulos, "Architecture for Web-based Services Integration," *The 29th Annual Conference of the Industrial Electronics Society (IECON '03)*, vol. 1, pp. 866-871, 2003.
- [7] B. Lee, "Embedded Internet Systems," *IEEE Internet Computing*, vol. 2, no. 3, pp. 24-29, 1998.
- [8] M. Ouzzani and A. Bouguttaya, "Efficient Access to Web Services," *IEEE Internet Computing*, vol. 8, no.2, pp. 33-34, Mar.-Apr. 2004.
- [9] J. Roy and A. Ramanujan, "Understanding Web Services," *IT Professional*, vol. 3, no. 6, pp. 69-73, Nov.-Dec. 2001.
- [10] U. Saif, D. Gordon, and D. J. Greaves, "Internet Access to a Home Area Network," *IEEE Internet Computing*, vol. 5, no.1, pp. 54-63, Jan./Feb. 2001.
- [11] J. W. Szymanski, "Embedded Internet Technology in Process Control Devices," *the Proceedings of IEEE International Workshop on Factory Communication Systems*, pp. 301-308, 2000.
- [12] W. T. Tsai, R. Paul, W. Song, and Z. Cao, "Coyote: An XML-based Framework for Web Service Testing," in the *Proceedings of 7th IEEE International Symposium on High Assurance Systems Engineering*, pp. 173-174, Oct. 2002.
- [13] J. P. Thomas, M. Thomas, and G. Ghinea, "Modeling of Web Services Flow," in the *Proceedings of IEEE International Conference on E-Commerce*, pp. 391-398, June 2003.
- [14] E. Topalis, G. Orphanos, S. Koubias, and G. Papadopoulos, "A Generic Network Management Architecture Targeted to Support Home Automation Networks and Home Internet Connectivity," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 1, pp. 44-51, Feb. 2000.

log_date	sender	txt_data	receiver
2004/07/28 14:01:10.89	002	地震狀況1	server
2004/07/28 14:01:09.81	002	地震狀況2	server
2004/07/28 14:01:08.78	002	地震狀況1	server
2004/07/28 14:01:07.76	002	地震狀況2	server
2004/07/28 14:01:06.68	002	地震狀況1	server
2004/07/28 14:01:05.67	002	地震狀況2	server
2004/07/28 14:01:04.64	002	地震狀況1	server

Fig. 6. The status of appliances by IE browser from database

Trust, Relationship Commitment and Supply Chain Integrations in China

Baofeng Huo, Xiande Zhao, Jeff Hoi Yan Yeung

Department of Decision Sciences and Managerial Economics, Faculty of Business Administration,
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China
{baofeng, xiande, jeff}@baf.msmail.cuhk.edu.hk

ABSTRACT

Supply chain integration (SCI) among different trading partners within a supply chain (SC) has received increasing attention from academicians and practitioners in recent years. However, our knowledge of what influences or enables SCI between trading partners, and how integrations in the SC influence the performance of the SC and firms within the SC, is still very limited. In this paper, we develop and test the measurement instruments for SCI and performance. We also investigate the impact of trust, relationship commitment on SCI and the impact of SCI on SC performance. We propose and empirically tested a model of SCI using data collected from manufacturers within the SC from China. The results show that trust and perceived strategic importance of relationship with the supplier and customers significantly improve relationship commitment. Relationship commitment positively influences SCI and SCI leads to SC performance and the financial performance of the firm within the SC. This study provides important insights for future researchers to understand trust, relationship commitment and SCI from various perspectives.

Keywords: Trust, Relationship Commitment, Supply Chain Integration, China

1. INTRODUCTION

Though the fundamental concept and importance of supply chain management (SCM) are widely accepted by both the scholars and the practitioners, there is a dearth of empirical research investigating how trust, relationship commitment influence supply chain integration (SCI) and how SCI influence the performance of the supply and the financial performance of firms within the supply chain. Previous studies identified different types of integration of SC (Markham, 2001, Narasimhan, 2001, Stank, Keller and Closs, 2001, Johnson, 1999, Morash & Clinton, 1998). Some papers analyzed the relationship between SCI and SC performance (Narasimhan, 2002, Armistead, 1993). Morgan & Hunt (1994) empirically validated the hypotheses of trust and relationship commitment as key mediating variables in ten forms of relationship marketing for tire retailers as well as their suppliers from the relationship marketing perspective. Their findings generalized that trust engenders cooperation among SC, and that relationship commitment is crucial for integrating SC partners into their key customers' business processes and established goals. Handfield & Bechtel (2002) demonstrated how suppliers and customers could build relationships based on trust and their findings further suggested that working to trust could improve SC responsiveness. Beth et al. (2003) advocated that trust and relationship commitment are placed in the highest priorities in achieving "SCI", a significant concept that promotes collaboration between SC partners for values and competitiveness.

Chen & Paulraj (2004) developed a theoretical framework that integrates strategic purchasing, supply management, logistics integration, supply network coordination, and SC performance. They made a good

attempt to synthesize a large body of literature. However, failed to examine how trust and relationship commitment enable SCI and the paper did not deal with the explanations of various types of SCI. Our study aims to build a model to represent the relationships among trust, relationship commitment, SCI, and SC performance using data collected from manufacturers from Mainland China and Hong Kong. Specifically, the objectives of the paper are as below:

- 1). To define three types of integration: customer integration, supplier integration, and internal integration and develop an instrument to measure them;
- 2). To study the relationships among these three types of integrations and how they influence SC performance and the financial performance of the firm within the SC; and
- 3). To investigate how trust and relationship commitment influence SCI.

2. LITERATURE REVIEW

Extensive literature has attached great importance to SCI for achieving competitive advantages (McGinnis & Kohn, 1993; Clinton & Closs, 1997), as well as operational performance (Ahmad & Schroeder, 2001; Frohlich & Westbrook, 2001; Stank, Keller and Closs, 2001). Several researchers such as Bowersox & Morash (1989) and Hammer (1990) suggested that SCI is to integrate the relationships, activities, functions, processes and locations among all channel members in the SC. Stevens (1989) proposed that integrating the SC is primarily involved in planning, coordinating and controlling materials, parts and finished goods from suppliers to customers at all different strategic, tactical and operational levels. Vickery et al., (2003) suggested that SCI should be strategically managed as a single

system as opposed to individually optimizing fragmented subsystems. But, one of the limitations of the earlier explanations is that they may not be systematic enough to differentiate integration from SCM. SCI could be illustrated as the degree to which the firm can strategically collaborate with their SC partners and collaboratively manage the intra- and inter-organization processes to achieve the effective and efficient flows of product and services, information, money and decisions with the objective of providing the maximum value to the customer at low cost and high speed (Bowersox, Closs & Stank, 1999; Towill & McCullen, 1999; Frohlich & Westbrook 2001; Vaart & Donk, 2003). Morash & Clinton (1998) investigated and compares two types SCI: external (customer and supplier) and internal (process reengineering) integration for approximately two thousand global firms. Markham (2001) investigated supplier and customer integration strategies in a global sample of 322 manufacturers. Stank, Keller & Daugherty (2001) developed and tested an instrument for measuring SCI competences as well as evaluating their relative importance to developing logistic distinctiveness.

Collaborative relationships have become one of the most significant resources for building up competitive advantage (Dyer & Singh 1998). Much literature (Kumar 1996; Spekman et al. 1998; Bachmann, 2001) argued that trust is a useful lubricant or fundamental ingredient in maintaining cooperation and avoiding conflicts. Handfield & Bechtel (2002) found that trust could significantly contribute to the long-term stability of a SC. Understanding trust has become the top priority concern in upholding the relationships among SC partners. Trust can be conceptualized in two different levels: interpersonal trust and organizational trust (Cummings & Bromiley, 1996; Zaheer et al., 1998). On the basis of interpersonal trust, organizational trust is established. Organizational trust can be further classified into two categories: intra-organizational trust (Shockley-Zalabak et al. 2000) and inter-organizational trust (Zaheer et al., 1998). Although the issue of trust within the SC has not been studied until recently, the concept of trust has drawn considerable attention in management (e.g., Kramer & Tyler, 1995; Doney, Cannon, & Mullen, 1998) and other disciplines (e.g., Child & Mollering, 2003).

Barber (1983) and Morgan & Hunt (1994) suggested that the propensity for relational continuity and the establishment of long-term relationship are primarily in the theme of "relationship commitment". Relationship commitment can be defined as the willingness of a party to invest resources into a relationship (Dion et al. 1992; Morgan & Hunt, 1994). Gundlach, et al. (1995) further pinpointed its importance for developing and sustaining successful relational exchange. Mayer et al. (1995) and Luo (2002) also argued that the commitment is the backbone of strengthening the function of established trust. Relationship commitment can be identified into two levels: interpersonal commitment and organizational commitment (Hornby, 1995). Organizational

commitment could be further categorized into Intra-organizational (Porter et al., 1974; Mowday et al., 1982) and inter-organizational commitment (Cheng et al., 2004). With increased inter-organizational commitment, SC organizations would develop closer relationships with their SC partners; therefore, it enhances the implementation of SCI. There has been an increasing research emphasis on trust and relationship commitment in recent years.

Handfield & Bechtel (2002) studied the role of trust and relationship structure in improving SC responsiveness using data from North American manufacturing firms. They suggested that trust can improve SC responsiveness and developing trust within the relationship improves supplier responsiveness even if buyers do not possess great control over their suppliers.

As cited by Chen & Paulraj (2004), a common measure of business performance is financial performance because the primary goal of business organization is to make profits for the shareholders. Financial performance has been widely used as a key measure of firm performance (Boyer et al, 1997; Boyer, 1999) and is evaluated in different dimensions. However, much literature (e.g. Dixon et al., 1990; Eccles & Pyburn, 1992) has pinpointed the limitations in relying solely on financial performance measures in SC. A broader conceptualization of performance measures includes customer service and other operational indicators. Neely et al. (1995) presented a few of the categories of performance comprising quality, time, flexibility, and cost. Van Hoek (1998) further advocated the SC firms to devise innovative measurement system as opposed to the traditional ROI-based system. Beamon (1999) presented an overview and evaluation of the performance measures used in SC models and also presents a framework for the selection of performance measurement systems for manufacturing SC. Vickery et al. (2003) included the dimensions of service performance in their customer service construct. Several researchers (Kaplan & Norton, 1992; Maskell, 1991; Gunasekaran, Patel & McGaughey, 2003) suggested that a balanced approach for the performance measurements is essential to present a clear picture of organizational performance. Actually, some recent SCI studies (Tan et al., 1998; Vickery et al., 2003) have used both operational and financial performances as indicators for the organizational performance. However, many SCI studies have measured either operational (Scannell et al., 2000; Stank, Keller and Closs, 2001) or financial performance outcomes (Ross, 2002).

3. THE PROPOSED MODEL AND HYPOTHESES

Based on an extensive literature search and in depth interviews with more than 15 practitioners who are in charge of SCM in Hong Kong and Mainland China, we propose the following theoretical framework for SCI (Figure 1). Within this framework, we have included the

following eight theoretical constructs, (1) Trust (trust with customer (Ctrust) and trust with supplier (Strust)). Trust is defined as a willingness to rely on exchange partner in whom one party has confidence (Moorman, et al, 1992). (2) Relationship Commitment (to customers (Rcc) and to suppliers (Rcs)). Relationship commitment is defined as the enduring desire of committed party to sustain an important on-going relationship (Morgan & Hunt, 1994). (3) Supplier Integration (Si). Supplier integration refers to the core competence derived from better coordination of all the critical suppliers in a company's SC to jointly achieve improved service capabilities at lower total SC cost (Bowersox, Closs, and Stank, 1999; Lee 2000). (4). Customer Integration (Ci), Customer integration refers to the core competence derived from better coordination of all the critical customers in a company's SC to jointly achieve improved service capabilities at lower total SC cost (modified from Bowersox, Closs, & Stank, 1999). Customer integration and supplier integration can be looked as external integration compared with internal integration. (5). Internal Integration (Ii). Internal integration refers to the degree to which a firm can structure its organizational practices, procedures and behaviors into collaborative, synchronized and manageable processes in order to fulfill the customer requirement. (Leveraged from Kingman-Brundage et al. 1995; Cespedes 1996; Kahn & Mentzer; 1996; Chen & Paulraj, 2004). (6) Perceived Strategic Importance of Relationship Commitment (for customer (Psicrc) and supplier (Psisrc)), (7) SC Performance (Scperf). SC performance is defined as the non-financial measures to evaluate the firm's customer responsiveness, such as SC cost, inventory level, quality, flexibility, lead-time, delivery, process time and speed, etc. (8) Company Financial Performance (Fperf). Company Financial performance is defined as financial and market measures to evaluate the firm's efficiency and effectiveness.

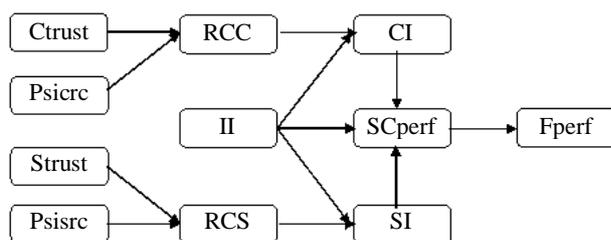


Figure 1. Proposed model

Trust and relationship commitment are salient and inter-related elements in structuring the inter-organizational interactions (Morgan & Hunt 1994) & Moore (1998). Based on the assumptions that the existence of trust reduces the perception of risk associated with opportunistic behavior and thus increases the exchange partners' confidence in the effectiveness of future relational exchange, motivating the exchange partners to commit in the relationship, Morgan and Hunt (1994), Moore (1998) and Ruyter et al. (2001) suggested that trust has an influence on relationship commitment. So we proposed the following hypotheses:

- H1: Ctrust has a positive influence on Rcc.
 H2: Psicrc has a positive influence on Rcc.
 H3: Strust has a positive influence on Rcs.
 H4: Psisrc has a positive influence on Rcs.

With relationship commitment, SC partners become integrated into their key customers' business processes and more tied to established goals (Morgan & Hunt, 1994; Chen & Paulraj, 2004). Johnson (1999) investigated the strategic role of inter-firm relationships through the concept of strategic integration. Stank, Keller and Daugherty (2001) developed and tested the measures to examine empirically the relationships between internal and external SC collaboration and logistical performance. Daugherty (2002) empirically examined the role of relationship commitment based on closer buyer-seller relationships. So we proposed that:

- H5: Rcc has a positive influence on Ci.
 H6: Rcs has a positive influence on Si.
 H7: Ii has a positive influence on Ci.
 H8: Ii has a positive influence on Si.

The relationship between SCI and performance outcomes is discussed over a vast body of SC and operations literature (Colin & Mapes, 1993; Daugherty et al., 1996 and Tan et al., 1998). For example, Armstead & Mapes (1993) identified the extent to which greater integration along the SC leads to improve operating performance. Stank, Keller and Closs (2001) developed and tested an instrument for measuring SCI competences as well as determining their relative importance to developing logistics distinctiveness. Stank, Keller & Daugherty (2001) suggested that collaboration with SC partners facilitates internal collaboration, which in turn enhances logistics performance. Frohlich et al (2001) demonstrated that SC companies with the widest degree of the arcs of integration achieve the highest level of performance improvement. So we proposed the following hypotheses:

- H9: Ci has a positive influence on Scperf.
 H10: Si has a positive influence on Scperf.
 H11: Ii has a positive influence on Scperf.
 H12: Scperf has a positive influence on Fperf.

4. RESEARCH METHODOLOGY

4.1 Sampling and Data Collection

To test the above hypotheses, we collected data from manufacturing companies in Hong Kong and Mainland China.

We selected five cities to represent China: Chongqing, Tianjin, Guangzhou, Shanghai, and Hong Kong. A mail survey combined with telephone calls was used to maximize the return rate. The questionnaire included questions on the demographic profile of the company and the questions related to the company's performance and SC performance, internal process integration,

customer integration and supplier integration, customer and supplier relationship commitments, and customer trust and supplier trust. For all these questions, a Likert scale of 1 to 7 was used. Before we launch the full-scale study, we piloted test the questionnaire using a sample of 15 companies. We revised the questionnaire based on the results of the pilot-test. We used the yellow pages of China Telecom in each one of the four cities in Mainland China and Directory of the Chinese Manufacturers Association in Hong Kong as a large sampling pool. We randomly selected the sampled companies from the lists and made telephone calls to these companies. Through the phone contacts, we make first make sure the company is a manufacturing company and then identify the right informant to fill out the questionnaire. We asked for the person who is in charge of supply chain management or the person who is knowledgeable about customer/supplier relationships and internal integration. We found the most suitable respondent is Supply Chain Manager, President, Vice President, or General Manager depending on the company. We get the company's permission to receive the questionnaire and name and address of the most suitable respondent for the survey. Then we sent the questionnaire to the most suitable respondent. A cover letter highlighted the objectives of the survey and its potential contributions to the respondents. Respondents were encouraged to participate in the survey with an entitlement to a summary report of the results. Self-addressed envelopes with return postage were also included together with the survey to facilitate the returning of the completed questionnaires. 4,569 companies were conducted via the phone and were identified to be manufacturers. A total of 1,356 agreed to receive the questionnaire and thus received the questionnaires. The response rate is 13.5% based on total manufacturing companies contacted and 45.5% based on the number of companies who received the questionnaire.

4.2 Measures, Construct Validity and Reliability

In the study, we use structural equation modeling to estimate the causal relationships among the different constructs with linear structural relations (LISREL) program and a sample of 617 companies. LISREL 8.54 was used to analyze the hypothesized model. A two-step model building approach was used, wherein the measurement models were tested for the reliability and validity prior to testing the structural model. For the measurement models, Trust with customers construct is measured by two items adapted from the scale developed by Morgan & Hunt (1994). Perceived strategic importance of customer relationship and perceived strategic importance of supplier relationship are measured by four items respectively. We modified the items of strategic integration used by Johnson (1999) to assess strategic importance on SCI. Relationship commitment with customer and relationship commitment with supplier are measured by three items respectively.

We adapted the scale developed by Morgan & Hunt (1994). Customer integration, supplier integration, and internal integration are measured by eight items respectively. The scales were largely derived from the items of Narasimhan & Kim (2002), Morash & Clinton (1998), and Narasimhan & Kim (2002). SC performance is measured by seven items that were partly derived from the items of Stank, Keller & Daugherty (2001). Financial performance is measured by five items that were partly derived from the items of "firm performance" in Narasimhan & Kim (2002).

Reliability is an assessment of the degree of consistency between multiple measurements of a variable (Hair *et al.*, 1998). The most commonly used measure of reliability is internal consistency. Flynn *et al.* (1990) suggested that the most accepted measure of a measure's internal consistency is Cronbach's alpha. The generally agreed lower limit for Cronbach's alpha is 0.50 to 0.60 (Flynn *et al.* 1990; Nunnally, 1994). Cronbach's Alpha tests were performed on the constructs. Since all alpha values are above 0.8, the items tested were deemed reliable. The data collected by surveys and other empirical designs is of little use unless its reliability and validity can be demonstrated (Flynn *et al.* 1990). There are two dimensions of construct validity: discriminant validity and convergent validity. Discriminant validity is the degree to which measures of different latent variables are unique (O'Leary-Kelly & Vokurka, 1998). Convergent validity relates to the degree to which multiple methods of measuring a variable provide the same results (O'Leary-Kelly & Vokurka, 1998). In our study, we try to assess the two types of validity by confirmatory factor analysis (CFA) in structural equation models. O'Leary-Kelly & Vokurka (1998) suggested that the CFA method of assessing convergent and discriminant validity is a more powerful tool and requires fewer assumptions than the traditional MTMM matrix method. At the beginning, we construct the CFA model using Lisrel program. In the model, each item is linked to its corresponding construct and the covariances among those constructs are freely estimated. The model fit indices are Chi-Square = 5197.43 with Degrees of Freedom = 1322, RMSEA=0.069, which indicate that the model is acceptable (Hu & Bentler, 1992). Generally, a construct with either loadings of indicators of at least 0.5, a significant t-value ($t > 2.0$), or both, is considered to be convergent valid (Fornell & Larcker, 1981; Chau, 1997). For our model, all of the factor loadings are greater than 0.50 and the t-values are all greater than 2.0. Therefore, convergent validity is achieved in our study. In order to assess the discriminant validity, we build a constrained CFA model, in which the correlations among constructs are fixed to 1. This model will be compared with the original unconstrained model, in which the correlations among constructs are freely estimated. A significant difference of the Chi-square statistics between the fixed and unconstrained models indicates high discriminant validity (Fornell & Larcker, 1981; Chau 1997). In our study, the differences of chi-square are significant at

0.001 level. So, the discriminant validity is ensured.

4.4 Structural Model and Hypotheses Testing

The structural model was analyzed based on the measurement models using the maximum likelihood estimation method. Table 1 shows the structural equation model and standardized coefficients. Eleven coefficients shown were significant at 0.05 levels. The initial model as shown in Figure 1 was tested, resulting in eleven significant path coefficients (H1-H9, H11-H12), suggesting the support in the data for the relationships. The data supported hypotheses H1-H9, and H11-H12, But hypothesis ten is not supported by the data. That means that supplier integration has no significant positive influence on SC performance. The results support the proposed structural equation model and the underlying theory except one hypothesis. The goodness of fitness indices for our model are $\chi^2(1355) = 5900.35$, NFI = 0.99, CFI = 0.99, RMSEA = 0.074, Standardized RMR = 0.015. These indices are better than the threshold values suggested by Hu & Bentler (1992). Therefore, our model can be accepted for future discussion.

Table 1. Standardized path coefficients

	Trust	Structure	Psychic	Process	Rcc	Rcs	Li	Si	Ci	Scperf
Rcc	0.17		0.83							
Rcs		0.15		0.85						
Ci					0.64		0.36			
Si						0.35	0.62			
Scperf							0.11	-	0.88	
Fperf										0.96

“-” means insignificant path coefficient at 0.05 level.

5. DISCUSSION AND MANAGERIAL IMPLICATIONS

The result of our hypothesis testing shows that a manufacturer’s trust with the supplier (customer) has significant positive impact on the relationship commitment to the supplier (customer). That means that a company with a higher level of trust with customers/suppliers is more likely to have a stronger commitment to its relationships with its customers/suppliers. To enhance relationship commitment with its customer or supplier, the company can either select partners that it can trust or build the trust through proper management of the relationships. The existence of trust can help to reduce the perceived risks in the relationships and thus enhance the relationship commitment. Based on trust, commitment is formed to extend the relationships and trust often precedes the committed investment. This conclusion was also supported by some other researchers. Morgan & Hunt (1994) and Moore (1998) stated that trust and relationship commitment are salient and inter-related elements in structuring the inter-organizational interactions. Greyskens, Bernedict, Steenkamp & Scheer (1996) found that higher trust increases affective

commitment between to marketing channel members. The results of our study highlight the importance of trust between the manufacturer, its supplier and its customer.

The result of our study also indicates that perceived strategic importance of relationship with supplier or customer significantly impacted relationship commitment. That means that companies that perceive the greater strategic importance of developing and maintaining relationships with its supplier or customers are more likely to have a level of commitment with their supplier or customers. Many companies have considered strategic partnerships with its customer or supplier as strategic asset and have used this asset to gain competitive advantages in the market place. Some companies have developed customer relationship management (CRM) systems to help develop and maintain their relationships with their customers. Our findings indicate that companies will make strong commitment to the relationships with their partners when they consider such a relationship as of strategic important to them.

Relationship commitment is found to significantly impact SC external integration. That means that companies with a stronger relationship commitment to customers/suppliers are more likely to have a greater extent of customer/supplier integration. Since SC integration requires investment in time and other resources, the partners must have strong commitment in the relationship before they are willing to invest in SC integration activities. With a higher level of relationship commitment, the customers or the suppliers are more likely to cooperate with the manufacturer. So the manufacture will have less difficulty to integrate the customers or suppliers with its own operation process in the SC to achieve the competitive advantages.

This finding is partially supported by the results of several previous studies in different contexts. Morgan & Hunt, (1994) and Chen & Paulaj (2004) found that SC partners become integrated into their key customers’ business processes and more tied to established goals with higher relationship commitment. Prahinski & Benton (2004) indicated that the buying firms should increase their efforts on cooperation and commitment in order to augment the supplier’s commitment to ensure better supplier performance. However, no previous studies have investigated the impact of relationship commitment on the SCI. Our results indicate that a company needs to gain commitment to the relationship with its supplier or customers in order to achieve a higher degree of integrations with them. To achieve a higher level of commitment, the company must first recognize the strategic importance of relationship with its supplier or customer. Furthermore, the company must develop trust with its customers or the suppliers. SC internal integration is found to have a positive influence on customer and supplier integrations. This means that companies with a greater extent of internal

integration are more likely to have a greater extent of integrations with its customer and supplier. The managerial implication of this finding is that the company needs to first integrate internal processes before it can effectively integrate with its customers and suppliers. This finding is different from what was found by Stank, Keller and Daugherty (2001). They suggested that collaboration with external SC entities increased internal collaboration, which in turn improves service performance. However in their study, they did not measure customer and supplier integration separately. Furthermore, the measurement items used in their internal and external collaborations do not capture all aspects of internal and external integrations. By using a more detailed list of measurement items, we are able to measure the different aspects of supply chain integrations. We believe that our findings are more logical. Internal integration is about the integration of the different internal functional departments, such as data integration, enterprise application integration, inventory management integration, periodic interdepartmental meeting, cross-functional teams, etc. The companies with a higher level internal integration can more effectively integrate their processes with the customers' or the suppliers' processes. If the companies have a bad internal integration, such as a low-level data integration, it is difficult for the companies to share or exchange information with their customers or suppliers. Without an integrative inventory management, a company will have little chance to share their production plans with the customers or suppliers. Therefore it makes more logical sense for a company to focus on internal integration first, then to integrate with the supplier and the customer.

The result of our study also indicate that both internal integration and customer integration significantly impacted SC performance, and SC performance has a positive influence on financial performance of the firm. It is interesting that supplier integration does not have a significant effect on SC performance. This might be because that manufactures in Mainland China and Hong Kong paid more attention to the customers. In order to get orders from the customers, manufacturers have to work hard to meet their requirements and respond to their requests quickly. Therefore, they are working hard to integrate their processes with the customers. As a result, customer integration contributed significantly to SC performance. For supplier integration, however, the manufacturer may not think that it is very important, and does not spend much resource to improve the inter-organizational processes with the supplier. They may just require the supplier to hold more inventories to buffer against uncertainty in the supply. This might be the major reason for the insignificant impact on the supplier integration on SC performance.

The relationship between SCI and performance has been discussed over a vast body of SC and operations literature. For example, Stank, Keller and Daugherty (2001) suggested that collaboration with SC partners

facilitates internal collaboration, which in turn enhances logistics performance. Frohlich & Westbrook (2001) demonstrated that SC companies with the widest degree of the arcs of integration achieve the highest level of performance improvement. Such previous researchers only tested the association between integration and SC performance. However, business executives may be concerned about both SC and financial performance. Our study also indicates that SC performance has a strong positive effect on financial performance. Therefore, manufacturer that want to improve financial performance needs to invest resources in both internal and customer integration to enhance the performance of the supply chain.

6. CONCLUSIONS

In this paper, we identified three types of integrations in supply chain and developed an instrument to measure SCI. We also investigated the relationships between the different types SCI, SC performance and the performance of the firm within the SC. Our result shows that internal integration positively influences supplier integration and customer integration. Furthermore, we found that customer integration and internal integration significantly influence SC performance and SC performance significantly influences the financial performance of the firm within the supply chain. However, supplier integration does not significantly influence SC performance. These findings enhance our understanding of how SCI influence performance and provide important guidelines for SC managers to enhance their financial performance through SCI.

In this study, we also investigated the impact of trust and relationship commitment on SCI. It was found that trust and perceived strategic importance of relationships with the customer and supplier significantly enhances relationship commitment with the supplier and customer. Relationship commitment, in turn, significantly improves SCI. Therefore, a company that wants to improve supply chain performance through SCI really needs to attach strategic importance to the relationship with the supplier and customer. Furthermore they must build trusting relationship with the customer and supplier.

This study is one of the first studies that examine the impact of trust and relationship commitment on SCI in from China. The findings from this study provide significant insights for practicing managers to enhance the performance of their supply chains. Furthermore, the models and instrument developed in this study can be used in future studies and thus this study also contributes greatly to the SCM literature.

Note: References are not provided due to space limitations. They are available from the authors upon request.

A Comparison of Brand Loyalty in Online and Physical Channels: An Empirical Application to Supermarkets

Javier Cebollada

Department of Management, Public University of Navarre, 31006 Pamplona, Spain,
cebollada@unavarra.es

ABSTRACT

In this study some differences in brand value in online channels are examined and compared with physical channels. The arguments presented lead to two hypotheses about brand loyalty in these environments. The first is that brand loyalty will be greater in online channels and the other that this difference is greater in product categories in which sensory factors predominate (those perceived by the senses, such as touch or smell). A choice model that depends on the degree of brand loyalty is proposed (Colombo and Morrison, 1989) and the hypotheses are contrasted empirically. Using data from 2003, the degree of brand loyalty and the degree of loyalty within a category are estimated for two product categories sold in an online and physical stores of one of the main Spanish supermarket chains.

KEYWORDS: brand loyalty, e-commerce, brand value, choice models, information availability and brand choice

1. INTRODUCTION

In this paper we propose to study some aspects of brand value in virtual or *online* commercial channels. Ever since the Internet began to be used for commercial purposes there has been a great deal of interest in ascertaining what differences exist between the online and traditional way of commercialising products. At first, many thought that the internet would cause an increase in competition that would lead to almost perfect competition, due to the increase in the size of markets, greater availability of information and the reduction in search costs for consumers. However, several studies soon showed that the Internet was a far from frictionless market. For instance, Brynjolfsson and Smith (2000a and 2000b) analysed prices of CDs and books in online and physical channels in purchases carried out using *shopbots* (web sites that compare prices of products in different online stores) and found that prices in the internet are often higher than those in physical stores, and that the most expensive internet stores were, at the same time, those with the greatest market share (e.g. Amazon). The explanation for these facts, according to the authors, is that consumers also use the brand as a sign of reliability and reputation and that considerable consumer loyalty exists on the Internet.

Subsequently, in various studies, the effect of brand and other marketing variables on consumer choice in the Internet have been studied. Degeratu et al (2000) compare consumer behaviour in an online store and a physical one. Using purchase data from a panel of consumers in several physical supermarkets and purchase data collected from an online supermarket's web site, the effect of brand, price and promotions are compared in the two environments for three product categories. In their analysis they include an interesting aspect that allows us to explain why the differences

found between the two environments depend on the product category: the role of the information available to the consumer in the brand choice process. The idea is that when sensory attributes predominate in a product category (attributes which are only perceived by the senses, such as smell or touch) the quality of information is lower in an online store and the effect of the brand is greater. On the other hand, when non-sensory attributes predominate in a product category (attributes which can be transmitted as information, for example the fat content of a margarine) the quality of information is higher in an online store and the effect of the brand is lower.

The role of information in consumer choice is an important aspect. In two articles by Alba et al (1997) and Burke et al (1992), it is seen that when consumers have less information (about the characteristics of a product) to take a decision, two things usually happen. Firstly, they grant importance to the brand (the brand, in these cases, plays a role which could be called the "halo effect" in the sense that it allows the value of the unknown attributes to be inferred). Secondly, consumers are much more likely to buy the same brand as on previous occasions, that is to say, they show greater brand loyalty. So, in situations of less information and more perceived risk, consumers minimise this by increasing loyalty. Generally, for the majority of product categories (at least those sold in supermarkets), it is more difficult to transmit information in an online store than in a physical store.

There are also other reasons for greater loyalty on the Internet. One is the fact that Internet consumers value highly the convenience and speed with which they can carry out their purchases (AIMC 2002), and so they are not willing to dedicate much time to these. One of the ways of simplifying and shortening the purchasing

process is to buy the same product as on previous occasions.

Another reason (closely related to the above) is the fact that tools exist on the Internet that allow the automatisation of purchases and may increase loyalty. In the case of online supermarkets, there are usually "personal lists" or "last shopping lists" that consumers can use in order to shop almost automatically, buying always the same products.

Therefore, in summary, our intention is to investigate two questions. Firstly, to see whether loyalty is greater in purchases carried out in online channels than in physical channels. As we have seen, there are sufficient arguments in favour of this position.

Secondly, we aim to verify whether, in product categories where sensory factors predominate, the difference in loyalty in online and physical channels is greater than in categories where sensory factors are less important. This, in a way, is equivalent to showing that when there is less information in the choice process, consumers are more loyal.

In order to examine both hypotheses concerning loyalty in online and physical environments, we will use a model that explains consumer behaviour as a choice problem between several discrete alternatives, and in which loyalty plays an important role. This is the Colombo and Morrison model (1989) which is described below.

We will subsequently carry out empirical analysis using purchase data from an online supermarket and a physical supermarket. We will estimate the parameters of the model for these data to see whether they support the aforementioned hypotheses. Finally, to conclude the paper, we will mention some of the study's principal limitations and some areas for future research.

2. A BRAND CHOICE MODEL WITH BRAND LOYALTY (COLOMBO AND MORRISON, 1989)

Let us take an individual who is faced with the choice of one alternative from a possible group of N (henceforth brands). Let A_1 be his previous purchase and A_2 his current purchase. The probability that his current purchase will be j , given that his previous purchase was i ($i, j=1, \dots, N$) is $P(A_2=j/A_1=i)=P_{ij}$ so that

$$P_{ij} = \exp U_{ij} / \sum_{k=1, \dots, N} \exp U_{ik}, \quad i, j=1, \dots, N \quad (1)$$

where $U_{ij}=U(A_2=j/A_1=i)$, is the utility of an individual currently purchasing j given that he previously bought i . Let us suppose that $U_{i,k}=U_k$ when $i \neq k$, that is, that the utility to the individual of the current purchase of a different product to the one bought previously is independent of this.

We define the "free" probability of purchasing i (without considering the previous purchase) as $P(A_2=i)=\pi_i$, $i=1, \dots, N$, where

$$\pi_i = \exp U_i / \sum_{k=1, \dots, N} \exp U_k \quad (2)$$

We now define α_i ($i=1, \dots, N$) as

$$\alpha_i = (\exp U_{i,i} - \exp U_i) / (\exp U_k + \exp U_{i,i} - \exp U_i) \quad (3)$$

We should observe that α may be considered as the degree of loyalty in brand i because it is monotonous in $U_{i,i}-U_i$, that is, the greater the influence of having bought brand i previously on buying it now ($U_{i,i}$ is great), the greater are $U_{i,i}-U_i$, and $\exp U_{i,i}-\exp U_i$, and viceversa. If we suppose that the individuals who previously bought i can be grouped in two segments, A and B , where with $U_{i,i}^A=U_i$ y $U_{i,i}^B \rightarrow \infty$, then we obtain $\alpha_i^A=0$ and $\alpha_i^B=1$.

If we now apply the above expressions, we can now rewrite the probabilities $P_{i,i}$ and $P_{i,j}$ as

$$P_{i,i} = \alpha_i + (1-\alpha_i)\pi_i \quad (4)$$

$$P_{i,j} = (1-\alpha_i)\pi_j \quad (5)$$

We may consider the α and π as the proportions of loyal and non-loyal customers to each brand.

On the other hand, if S_i is brand i 's market share, estimated from the first purchase, then $LP = \sum_{i=1}^N \alpha_i S_i$ is the degree of loyalty in the market.

Estimation of the parameters α and π is carried out using the maximum likelihood method.

3. EMPIRICAL ANALYSIS

3.1. Data

We are going to conduct an empirical application of the above model in order to verify whether it supports the two hypotheses mentioned in Section 1. We will use data from one of the five leading supermarket chains in Spain, present in a large part of the country, and a pioneer in Internet sales. For reasons of confidentiality, we will omit the name of the chain.

We will use data corresponding to purchases from June to November 2003 obtained at the point of sale in the physical supermarkets and at the web site of the online store, for consumers resident in several regions of the country, principally in large towns, which is where this distribution chain is established. Internet purchases are possible in practically the same towns as those in which physical supermarkets exist. The variety and depth of product lines are similar in both types of supermarket, as are the prices and the majority of promotions.

We will use data from two product categories (at a later date we intend to generalise the study using a larger number of categories): kitchen paper and domestic aluminium foil. The reason for choosing these categories is that for one of them, kitchen paper, sensory factors that are best appreciated in the physical supermarket, such as, texture, design, thickness, size etc. predominate. In the case of the aluminium foil, however, few of these sensory attributes exist: the packaging is opaque, and so one cannot see inside, and length in metres is a non-sensory factor that can be transmitted easily via Internet.

3.2. Transition matrices

These matrices reflect the number of consumers who have bought on a certain occasion one brand and another (or the same) the next time. Each element $n_{i,j}$ is the number of consumers who purchased brand i the first time and brand j the second. From these elements we can estimate the degrees of loyalty (α) in each brand as well as in the category as a whole (LP). Tables 1 and 2 show the transition matrices for both categories in both stores.

3.3. Results of the estimation

Table 3 shows the results of the estimation of the loyalty parameters for the brands (α) and within the product category (LP) in each store and for both product categories.

It can be seen that category loyalty is greater in both categories in the online store than in the physical one (0.80 compared to 0.41 in kitchen paper and 0.96 compared to 0.79 in aluminium foil). This confirms the first hypothesis that loyalty is greater in online stores than in their physical counterparts.

Furthermore, the difference between the two store types is greater for kitchen paper, which, as we have seen, is the category with most sensory factors that are

difficult to communicate to consumers in an online store. This result confirms the second hypothesis.

4. LIMITATIONS AND FUTURE RESEARCH

At present this research is ongoing, and so there are a number of limitations and aspects to be analysed. One of the limitations is that we have analysed only two categories. In order to generalise the results we need to study a larger number of categories in which sensory and non-sensory factors clearly predominate.

It would also be interesting to analyse the role of prices and promotions. In the supermarket chain considered here these variables are similar, but they could produce different effects in the online store than in the physical one.

Another interesting aspect to analyse is how brand loyalty varies between brands, the model allows this to be done. For instance, we could compare private brands with leading brands (in the online store private brands enjoy the additional advantage of being able to be listed separately).

There may also be other factors in play apart from sensory and non-sensory attributes which can affect the transmission of information that influences consumer behaviour in both environments. For example, an interesting area to research is whether the differences between the two store types is due to the fact that they are used by different consumer types.

To conclude, in summary, we can state the following. We have analysed brand value through the degree of brand loyalty in online and physical stores and explained this in terms of the way information is transmitted in the two store types. We have related the presence of sensory factors with the online stores' difficulty in transmitting information. Finally, we have confirmed the two hypotheses put forward empirically.

Table 1. Transition matrices. Kitchen paper.

		Online supermarket			Physical supermarket					
		Second purchase occasion			Second purchase occasion					
		Private label	Colhogar	Scottex	Bols	Private label	Colhogar	Renova	Scottex	
First purchase occasion	Private label	70	6	0	Bols	3	10	5	0	0
	Colhogar	5	38	1	Private label	3	140	52	1	18
	Scottex	1	1	3	Colhogar	5	49	128	2	14
					Renova	0	1	1	1	1
					Scottex	2	10	18	2	25

Table 2. Transition matrices. Aluminium foil

Online supermarket				Physical supermarket			
		Second purchase occasion				Second purchase occasion	
		Albal	Private label			Albal	Private label
First purchase occasion	Albal	10	2	First purchase occasion	Albal	23	9
	Private label	0	39		Private label	10	136

Table 3. Results of the estimations

Kitchen paper							
Online store				Physical store			
Brand	Alfas	Pis	Market share	Alfas	Pis	Market share	
Private label	0.86	0.43	0.61	0.46	0.36	0.44	
Colhogar	0.73	0.50	0.35	0.42	0.39	0.40	
Scottex	0.57	0.07	0.04	0.32	0.17	0.12	
Bols				0.12	0.05	0.04	
Renova				0.23	0.03	0.01	
Degree of loyalty in the category		0.80		0.41			
Domestic aluminium foil							
Online store				Physical store			
Brand	Alfas	Pis	Market share	Alfas	Pis	Market share	
Private label	1.00	1	0.76	0.87	0.47	0.82	
Albal	0.83	0	0.24	0.41	0.53	0.18	
Degree of loyalty in the category		0.96		0.79			

Note. Coefficients are significant at 10%

REFERENCES

- [1] AIMC (2003) Navegantes en la red: Quinta encuesta AIMC a usuarios de internet (www.aimc.es)
- [2] Alba, J.W., Lynch, J., Weitz, B., Janiszewski, C., Lutz, R., Sawyer, y A., Wood, S., (1997). Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentives to Participate in Electronic Marketplaces. *Journal of Marketing* 61, 38–53.
- [3] Anderson, N.H., (1981). *Foundations of Information Integration Theory*. Academic Press, New York.
- [4] Bakos, J.Y. (1997). "Reducing Buyer Search Costs: Implications for Electronic Marketplaces," *Management Science*, 43 (December), pgs. 1676–1692
- [5] Bettman, J.R., y Kakkar, P., (1977). Effects of information presentation format on consumer information acquisition strategies. *Journal of Consumer Research* 3, pgs. 233–240.
- [6] Brynjolfsson, E. y Smith, M. (2000a) "Frictionless Commerce? A Comparison of Internet and Conventional Retailers", *Management Science*, April, 46-4, pgs 563-585.
- [7] Brynjolfsson, E. y Smith, M. (2000b) "The Great Equalizer?. Consumer Behavior at Internet Shopbots, MIT Working Paper
- [8] Burke, R.R., (1997). Do you see what I see? The future of virtual shopping. *Journal of the Academy of Marketing Science* 25, pgs. 352–360.
- [9] Burke, R.R., Harlam, B.A., Kahn, B., Lodish, L.M., (1992). Comparing dynamic consumer choice in real and computer-simulated environments. *Journal of Consumer Research* 19, pgs. 71–82.
- [10] Burke, S., (1992). The effects of missing information and inferences on decision processing and evaluation. Unpublished PhD thesis, The University of Michigan.
- [11] Colombo, R.A. y Morrison, D.G. (1989), "A Brand Switching Model with Implications for Marketing Strategies", *Marketing Science*, Vol. 8, N. 1, pgs. 89-99.
- [12] Degeratu A.M., Rangaswamy A. y Wu J. (2000), "Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes ", *International Journal of Research in marketing*, 17, pgs. 55-78
- [13] Hoffman, D.L., Novak, T.P., (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing* 60, pgs. 50–68.
- [14] Kamakura, W.A., Russell, G.J., (1993). Measuring brand value with scanner data. *International Journal of Research in Marketing* 10, pgs. 9–22.
- [15] Lynch Jr., J., Ariely D., (1998). Interactive home shopping: Effects of search cost for price and quality information on sensitivity, satisfaction with merchandise, and retention. Working Paper, Fuqua School of Business, Duke University.

A Defection Detection Procedure Using SOM and Markov Chain : A Case of On-Line Game Provider

Kyunghee Chae¹, Heeseok Song², Juchoel Choi¹

¹ School of Business Administration, KyungHee University, Seoul, 130-701, Korea

² Department of Management Information Systems, Hannam University, Taejon, 306-791, Korea
khchae@gmail.com, songorak@hanmir.com, hssong@hannam.ac.kr

ABSTRACT

Customer retention is very important issue. In this reasons, many studies have been conducted for customer retention to demonstrate the potential of data mining through experiments and case studies. But current customer retention procedures only focus on detection of potential defectors based on likelihood of defection overlooking cost. In this paper, we propose an economic procedure for profit maximizing based on the customers' LTV and behavior state transition cost. For the aim of this paper, we use past and current customer behavior by integrating data mining techniques. In this procedure, SOM is used to determine the possible states of customer behavior from past behavior data. Based on this state representation, a Markov chain is applied to generate likelihood of state transition.

Keywords: *Customer relationship management, Data mining, Life-time value, Markov chain, Self-organizing map*

1. INTRODUCTION

Customer retention and life-time value(LTV) are an important issues in today's competitive environment. It is mapping customer data to define customer behaviors so that the processes of a company are fully occupied in acquiring, selling to, and maintaining a long-term relationship to a customer. The longer a customer stays with a company the more profit the customer generates. This is the outcome of a number of factors such as the higher initial costs of attracting new customer, the increase in the number of purchases, the long-term customer's better understanding of the company, and positive word-of-mouth. In this reasons, many studies have been conducted for customer retention to demonstrate the potential of data mining through experiments and case studies. But current customer retention procedures only focus on detection of potential defectors based on likelihood of defection overlooking cost.

This paper proposes an economic procedure for profit maximizing based on the customers' LTV and behavior state transition cost. For the aim of this paper, we use past and current customer behavior by integrating data mining techniques. Here, the term 'customer behavior' means action-oriented activities like calling, visiting Websites, and making purchases. However, predicting and understanding of customers' behavior for multiple periods is not easy because of the following reasons. First, customer defection cannot be predicted by only one behavior feature. Second, these features may also be related to customer defection in a highly nonlinear way. For this reasons, SOM(Self-Organizing Map) is used. In this procedure, SOM is used to determine the possible states of customer behavior from past behavior data. Because it facilitates the understanding of complex behavior dynamics so that several variables and their interactions can be inspected simultaneously. Based on

this state representation, a Markov chain is applied to generate likelihood of state transition. Also to maximize company's profit, the proposed procedure recommends the desirable behavior state for the next period based on the economic analysis of transition cost and LTV. The transition cost is based on the likelihood of state transition because that low likelihood is meant to be hard to recommend a desirable behavior state, if so we have to make more effort. Here, the term 'desirable behavior state' means that it has high LTV.

Based on the state representations, the following procedures are developed: (1) collecting behavior data and pre-processing, (2) determining the behavior states through SOM training, (3) determining a desirable state, (4) developing a marketing strategy for profit maximizing. We applied our procedures to the online game provider, because the histories of customer behavior are easily collected during multiple periods. Although our procedures are explained for a specific online game company, they can be applied to various service industries with fluent customer behavior data such as telecommunications, internet access services, and contents providing services. We begin by reviewing the previous study on defection detection in section 2 and the concept of SOM and Markov chain, which are prerequisites for our research, are summarized. In section 3, our suggested procedures are explained step by step. A case study is presented to illustrate and evaluate our procedures in section 4. Finally we summarize our contributions and outline areas for further research in section 5.

2. RELATED WORK

2.1 Existing Works In Customer Retention

Many Studies [2,5,6,7] have been conducted to detect potential defectors and preserve customer using data

mining techniques. Logit regression, decision trees and neural networks are the most frequently used technique to predict the likelihood of defection in these studies. However, these studies use static model to predict the likelihood of defection in spite of continuous and dynamic property of individual access behavior. Song et al. [7] and Kim et al. [5] proposed a solution that use static model to predict the likelihood of defection in spite of continuous and dynamic property of individual access behavior, but it also overlooked cost that is very important factor to make a decision. Compared to these existing works, our research builds an economic procedure for customer retention. Not only our procedure could predict potential defector, but also give the guidance to control whose undesirable behavior which leads to defection.

2.2 SOM

SOM was developed in its present form by Kohonen [4] and thus they are also known as Kohonen Maps. The SOM is able to map a structured, high-dimensional data onto a much lower-dimensional array of neurons in an orderly fashion. The mapping tends to preserve the topological relationships of the input data. Topological preserving means that the data points lying near each other in the input space will be mapped onto nearby map units. Due to this topology preserving property, the SOM is able to cluster input information and their relationships on the map. In this reasons we adopt SOM to determine every possible behavior state for a specific domain. The SOM has proven to be a valuable tool in data mining, full-text mining and financial data analysis. It has also been successfully applied in various engineering applications in pattern recognition, image analysis, process monitoring and fault diagnosis [8].

2.3 Markov Chain

A Markov chain, which is one special type of discrete-time stochastic process, can be used to predict the evolution of customer states [3].

A Markov chain has the following properties and assumptions:

? The probability distribution of the state i at time $t \geq 1$ only depends on the state at time t .

? For all states i and j , $p(X_{t+1} = j | X_t = i)$ is independent of time t . These two assumptions can be specified by the conditional probabilities for a Markov chain:

$$p(X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, K, X_1 = i_1, X_0 = i_0) = p(X_{t+1} = j | X_t = i) = p_{ij}$$

for all states i_0, i_1, K, i_t, i, j

We adopted the probability distribution of the state i at time $t \geq 1$, and an absorbing Markov chain in which some of the states are absorbing and the rest are transient

ones. An absorbing state has a special transition probability, $p_{ij} = 1$ [9].

3. PROPOSED PROCEDURE

3.1 Overall Procedure

In this section, we present the overall procedure of our methodology for defection detection and prevention with Figure 1. As shown in Figure 1, the proposed procedure consists of four steps. In the first step, all the behavior data related to defection is collected from operational databases and Web log files. In the next step, the SOM model is developed to determine all the possible behavior states which are mutually exclusive and exhaustive. The SOM is used to represent the state of each customer behavior from collected behavior data. After determining the behavior states using the SOM, all the continuous and complex behavior data is converted into discrete states. Based on this state representation, we determine a desirable state that is low defection rate, low transition cost, and high LTV. In the final step, we make a campaign plan to focus or modifying their behavior patterns for profit maximizing.

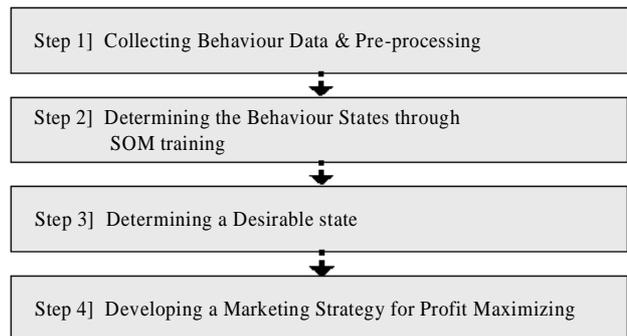


Figure 1. Overall procedure

3.2 Step 1: Collecting Behavior Data and Pre-processing

First of all, features that can affect defection are selected based on the prior knowledge of domain experts, and the value of features are collected over the multiple periods for each user from Web log files, purchasing databases, claim databases and the other sources. After collecting behavior data, data preparation is needed to build behavioral profiles. In particular the Web log data requires a great amount of effort to be transformed into usage data because of the difficulties in session identification and user identification. More detailed information for data preparation of the Web log can be found in Cooley et al.[1].

3.3 Step 2: Determining the Behavior States through SOM Training

Based on the selected input features, all the possible states of customer behavior are determined that are prerequisite for the SOM is a value array of the filtered

features such as the average length of session, the number of sessions in unit period, the number of congestions per unit period, and so on. In determining the behavior states using the SOM, the total number of states has to be given. If we set the total number of states to be very large, then the prediction model will lose the generalization capability because of overfitting. On the contrary, if we specify the total number of states to be very small, then the model may not have discrimination power between defectors and non-defectors. Therefore, choosing the number of states must be done carefully. The decision for determining the number of states is usually dependent on the characteristics of the domain and the goals of decision-makers.

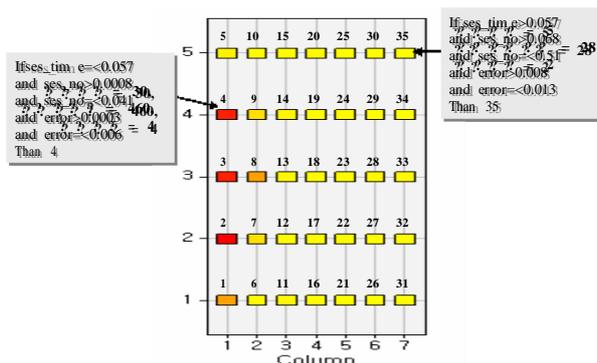


Figure 2. Behavior states and their interpretations

The satisfactory number of states may be given by trial and error using past multiple periods' behaviour data and the target variable (i.e., whether they actually defect or not). To evaluate the performance of each number of states, FalsePositive and FalseNegative are used as a measure in our procedure. FalseNegative error is more important than FalsePositive error because FalseNegative is more costly and dangerous than FalsePositive. If a non-defector is incorrectly classified as a defector (i.e., FalsePositive), the customer will receive a large concession, but if a defector is incorrectly classified as a non-defector (i.e., FalseNegative), the customer will terminate services before a company reaches the customer and thus the company will lose revenue [2]. Figure 2 illustrates the possible behavior states which are the results of SOM learning. To interpret each state on the map, a decision tree analysis(C5.0) is additionally conducted.

3.4 Step 3: Determining a Desirable State

In this step, we determine a desirable state for the aim of this paper that is the longer a customer stays. For determining a desirable state, we find a lower defection rate state at first. And then, we estimate profit rate of each state and recommend maximum profit state.

3.4.1 Average defection rate of each state

To determine a desirable state, at first, average defection rate have to be computed of each state : $Df_i = D / N$

Where, Df_i : the average defection rate of the state i
 D_i : the number of defection in the state i
 N_i : the total number of state i

3.4.2 Transition cost using Markov chain

The transition cost is based on the likelihood of state transition because that low likelihood is meant to be hard to recommend a desirable behavior state, if so we have to make more effort. The likelihood of state transition is generated from making an absorbing Markov chain. To make an absorbing Markov chain, we have two assumptions. First, we assume that customer behavior states have the Markovian property. It implies that given the present state of the customer, the future probabilistic behavior of the customer is independent of its past history and only depends on the current state. This is because the customer's current state information most impacts the decision of the customer's next behavior. Second, we assume that all the customers can be potential defectors because they eventually leave the company some time or other if they are not continuously satisfied with the company's service in today's greatly compressed online game industry.

Table 1. The transition probability matrix

		State at time t+1					
		1	2	3	4	---	def
State at time t	1	0	0	0.2 9	0.1 4	---	0
	2	0	0.0 8	0.2 3	0.0 8	---	0.0 8
	---	---	---	---	---	---	---
	34	0	0	0	0	---	0
	35	0	0	0	0	---	0.0 9
def	0	0	0	0	---	1.0 0	

To make a transition probability matrix, we used the states of all the customers for four periods (i.e., four weeks). We rearranged four pairs of transition data for each customer such as all of the transition data from t-4 week to t-3 week, from t-3 week to t-2 week, from t-2 week to t-1 week, and from t-1 week to t week (i.e., an absorbing state). And then we calculated transition cost, C_{ij} , based on a transition probability from state i to another state j as shown in Table 1.

3.4.3 LTV generation using an absorbing Markov chain

In the Table 1, state 'def' means 'defector' which is an absorbing state. According to the property of an absorbing Markov chain, if the customer's current state begins in a non-absorbing state (i.e., from state 1 to state 35), then the customer will leave the non-absorbing state

and end up in one of the absorbing states (i.e., state ‘def’) eventually [9]. Figure 3 describes the general transition matrix for an absorbing Markov chain. It consists of four parts which are I, O, R and Q, as illustrated in Figure 3. We can find out the elapsed time to absorbing state ‘def’ in each non-absorbing state by using the matrix $(I - Q)^{-1}R$.

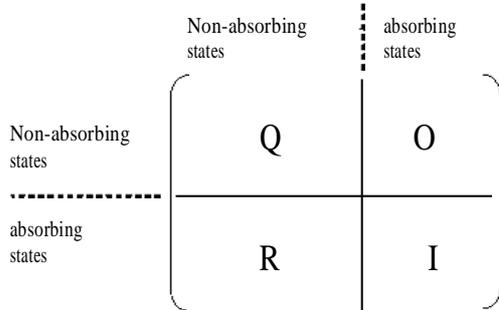


Figure 3. The transition matrix for elapsed time

Table 2 describes the outcome of the matrix $(I - Q)^{-1}R$. In this Table, we can see that if a customer belongs to state 1 at time t , then the customer will defect 11.5 periods later on average. The elapsed time to absorbing state ‘def’ can be used as LTV generation. We use simple LTV that only considered period (i.e., elapsed time) and profit (i.e., fee about session time): $LTV_i = E_i P_i$

Where, E_i : the elapsed time of the state i

P_i : the average profit of the state i

Table 2. Sample of the elapsed time to an absorbing state

State	Elapsed time (Week)	State	Elapsed time (Week)
1	11.5	6	14.1
2	9.8	7	12.9
3	7.5	8	9.4
4	11.3	9	10.5
5	8.2	10	12.2

3.4.4 Recommend a desirable state

In this step, we estimate gain of each state and recommend a desirable state that is maximum profit state. As it mentioned, we find a desirable state using a defection rate, and then desirable state is determined based on the profit that is transition cost and the LTV.:

$$Gain_{ij} = (LTV_j - LTV_i) / C$$

3.2 Step 4: Developing a Marketing Strategy

For the defection prevention, we will design segmental campaigns based on the information of a To-Be state.

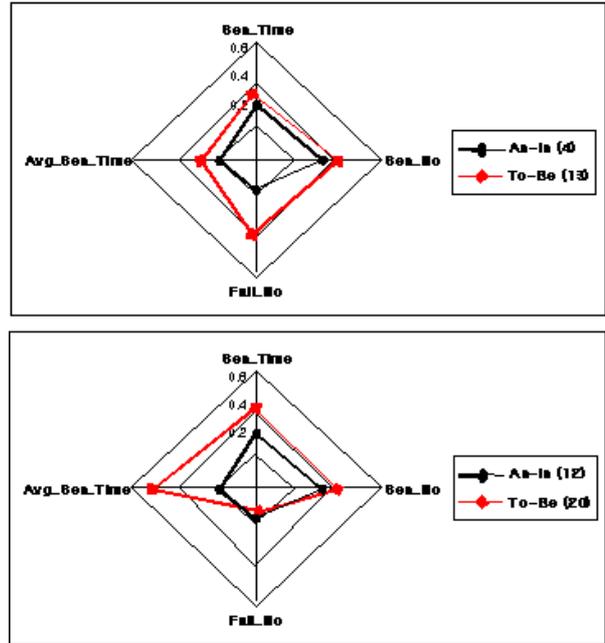


Figure 4. Segmental campaign design

Figure 4 provides samples of the campaign design. An As-Is behavior state and a To-Be behavior state for two segments are presented in this figure. From this figure, the campaign design for ‘4’ should focus on reduction error, and for ‘12’ should focus on extending session length. We will illustrate a case of campaign design in section 4.

4. A CASE STUDY

4.1 Data

The dataset for the case study is prepared from an online game provider in Korea. We sampled 255 customers and collected totally 114,736 transactions for those customers from various sources such as web log data and the transaction database. The collected input data for defection prediction contain the total session length for a week, the total number of sessions for a week, the total number of access errors caused by congestion, and the average length per session. Customer profiles were maintained for the four input features during four consecutive weeks for each customer with an actual defection indicator. We defined a defector as the customer who has no session time for one month, because there is no voluntary sign of withdrawal in this Internet game site.

4.2 Determining the Behavior States

As it mentioned, FalseNegative error is more important than FalsePositive error because if a non-defector is incorrectly classified as a defector (i.e., FalsePositive), the customer will receive a large concession, but if a defector is incorrectly classified as a non-defector (i.e., FalseNegative), the customer will terminate services

before a company reaches the customer and thus the company will lose revenue. Therefore, we made an experiment on several cases and select a minimum of FalsePositive error in FalseNegative error is limited to 5%. The outcome is described in Table 3 and ‘5 □7’ matrix, FalsePositive error is 17.6% and FalseNegative error is 4.6%, is appropriate to our case.

Table 3. FalsePositive and FalseNegative error matrix

Actual \ Predicted	6 □6		5 □7		7 □5	
	Def	Non	Def	Non	Def	Non
Defector	-	11.7	-	17.6	-	15.5
Non-defector	11.9	-	4.6	-	6.9	-

4.3 Determining a Desirable State

For determining a desirable state, we programmed a profit generation automatically using a Visual Basic 6.0. Table 4 provides desirable states(i.e., to-be) as a result and differences between as-is and to-be.

Table 4. Differences between as-is and to-be

As-is	To-be	Differences between as-is and to-be
1	2	Increase in session number
2	3	
3	10	
4	10	Increase in average session time
16	24	
17	18	
18	25	
9	10	Increase in session time
10	18	
11	12	
13	12	
22	25	
23	24	Increase in session number and time
12	25	
19	18	
21	25	Decrease congestion number
24	25	
27	25	Increase in session number and decrease congestion number
29	23	
32	24	

4.4 Developing a Marketing Strategy

Table 5 is a real cases and gives some idea about campaign object and feasible campaign based on Table 4. For instance, if we want to induce a customer in state ‘1’ to desirable state ‘2’, then we should focus on extending session number. And in this object, we can make a campaign plan like Table 5.

Table 5. Sample of the feasible campaign

Object	Feasible campaign
Extending session time	- Providing additional free access time - Providing gifts and awarding prizes based on the accumulation point
Extending session number	- To provide free coupon - Recommend new game information - Frequent update of the site and make an event
Reducing congestions	- Recommending the use of non-peak time by suggesting a discount rate - Sending a manual for congestion

4.5 Evaluation

In this section, we compare the prediction accuracy of the proposed procedure to that of the MLP neural network and a decision tree. To evaluate the performance of our defection detection procedure, we use FalsePositive and FalseNegative as measures. Figure 5 shows the comparative results of prediction accuracy among our suggested model and the MLP neural network(MLP NN) based on same input data in our domain(online game provider). Our proposed procedure resulted in both a slightly higher recall and hit ratio. This means that our proposed procedure can cover defection prevention as well as defection detection without deterioration of prediction accuracy compared to that of the MLP neural network.

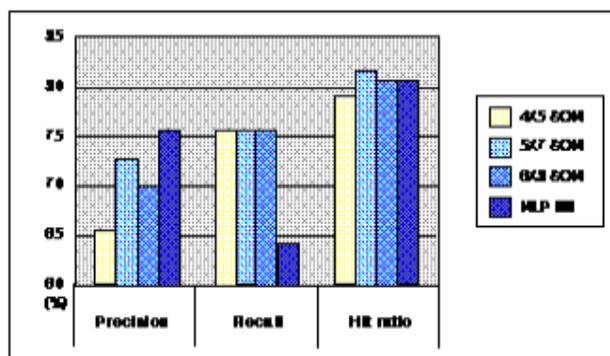


Figure 5. Comparative results of prediction accuracy

6. CONCLUSION

We proposed an economic procedure for profit maximizing based on the customers' LTV and behavior state transition cost. For this purpose, possible states of customer behavior are determined from past behavior data using SOM. Based on this state representation, the transition cost and LTV can be predicted using a Markov chain. In addition, our proposed procedure is extended to defection prevention for potential defectors and it assists in building segmental campaign plans by recommending the desirable behavior state for the next period to lower

the likelihood of defection. Our proposed approach is a practical implementation procedure of eCRM because it tries to maintain a relationship with potential customers using an automated campaign procedure continuously

As an area for further research, we have a plan to develop a system for defection detection and prevention based on our suggested procedure. In this study, we applied our procedure to the online game provider, but we believe that it can be applied to other service industries such as telecommunications, Internet access services, and contents providing services, and check the effectiveness of our proposed procedure.

REFERENCES

- [1] Cooley, R., B. Mobasher, J. Srivastava, "Data preparation for mining world wide web browsing patterns," *Knowledge and information systems*, Vol. 1, No. 1, pp5-32, 1999.
- [2] Datta, P., B. Masand, D. Mani, B. Li, "Automated cellular modeling and prediction on a large scale," *Artificial Intelligence Review*, 14, pp485-502, 2000.
- [3] Ha, S. H., S. M. Bea, S. C. Park, "Customer's time-variant purchase behavior and corresponding marketing strategies: An online retailer's case," *Computers&Industrial Engineering*, Vol. 43, pp801-820, 2002.
- [4] Kohonen, T., "The Self-Organizing Map," *Proceedings of the IEEE*, Vol. 78, No. 9, pp1464-1480, 1990.
- [5] Kim, Y. A., H. S. Song, S. H. Kim, "A dynamic procedure for defection detection and prevention based on SOM and a Markov chain," *International Conference of Korea Intelligent Information Systems Societ*, pp141-148, 2003.
- [6] Ng, K., H. Liu, "Customer retention via data mining," *Artificial Intelligence Review*, Vol. 14, No. 6, pp569-590, 2000.
- [7] Song, H. S., J. K. Kim, Y. H. Cho, S. H. Kim, "A personalized defection detection and prevention procedure based on the self-organizing map and association rule mining :Applied to online game sit," *Artificial Intelligence Review*, Vol. 21, No.2, pp161-18, 2004.
- [8] Vesanto, J., "SOM-based data visualization methods," *Intelligent data analysis*, pp111-126, 1999.
- [9] Winston, W. L., *Operations research applications and algorithms*, 3rd ed. Duxbury press, 1994.

A Global Comparative Study of Online Auction Markets

James K. Ho

Liautaud Graduate School of Business, University of Illinois at Chicago, IL 60607, USA
jimho@uic.edu

ABSTRACT

Online auctions have demonstrated how e-commerce can transform business, and not merely transplanting conventional processes to a new medium. The resulting emergence and growth of diverse markets pose the intriguing question of what 'shape' a given market is in at a particular moment of development. The approach of Topological Analysis, which is based on the Star Plot method in data visualization, provides such a model using only operational data, without any expert knowledge of the specific auction market, or financial details from the transactions. Using extensive analysis of data available on eBay.com, the foremost online auction platform to date, a global comparative study of four markets in eight countries is conducted. The results shed light on the evolution of electronic markets, as well as their regional and cultural variations.

Keywords: online auctions, market topology, data mining, data visualization

1. DATA VISUALIZATION AND THE SHAPE OF ONLINE AUCTION MARKETS

The Internet is a new medium of communication connecting potential partners in e-business worldwide. The initial frenzy over its promises led to grossly exaggerated valuations of business models that were mere transplantations of existing processes to the alternative channel. Now that the bubble has burst [8], more sensible and critical thoughts can be turned to true transformations that are creating and nurturing markets of the future, and shaping business strategies in a networked world. Online auction is one of the very few cases that has held a steady course ([5], [6], [7], [10]), as evidenced in the success to date of eBay.com. Founded in September 1995, eBay has become a global trading platform where on any given day, there are more than 16 million items listed across 27,000 categories. In 2003, at least 30 million people will buy and sell well over \$20 billion in merchandise, so that the entire culture it engenders is now being described as the eBay economy [3]. It is also the richest source of data for online auction markets, as records of all its transactions are available to the public (on a 2-week rolling basis). In (Ho 2004) the intriguing question of what 'shape' a given market is in at a particular moment of development was posed. Using the Star Plot method of high-dimensional data visualization [4], a topological model was developed, based only on operational data, without any expert knowledge of the specific auction market, or financial details from the transactions. The dimensions in this model are reviewed in the next section. Applications of this Topological Analysis include the study of market evolution over time, market design and simulation in experimental economics, as well as inter-brand or international comparative studies of online auction markets. As a full-scaled example of the latter, the results for four markets (digital cameras, classical CDs, diamond rings, Star Wars play-figures) in eight countries/regions (US, Canada, Australia, United Kingdom, France, Italy, Spain, Taiwan) are presented.

2. TOPOLOGICAL ANALYSIS OF ONLINE AUCTION MARKETS

In [2], a topological model for an online auction market is defined to be a simultaneous graphical display of all the dimensions of its relevant data base, which provides a geometrical shape as a descriptive, visual statistics of the market. In particular, various dimensions were identified from available data for constructing a specific topological model that can help discern market efficiency. Such a model provides a visual cue for whether a market is favorable to buyers or sellers, without expert knowledge of the items involved or the prices attained. The dimensions in this model are briefly reviewed below.

Gross Activity

The absolute size of a market is indicated by the average number of auctions being offered per day.

Net Activity

As a categorical measure of the effective activity in an auction market, the proportion of listings that ended with at least one bid is used. In cases where a reserve price is in effect, the listing is counted whether the reserve is eventually met or not.

Participation

For those auctions that resulted in at least one bid, an aggregate measure of activity is the average number of bids. Multiple regular bids from the same buyer are counted separately.

Seller Diversity

For a sense of whether a market is dominated by relatively few sellers, or there is diverse sources of supply, the number of auctions in the data collection offered by each individual seller is tracked and the distribution is used as an indicator of seller diversity

Seller Experience

At the completion of any auction on eBay, the buyer and seller have the opportunity to post feedback to the system. The net feedback rating is used as a surrogate measure of an eBay's experience. The distribution of four levels of seller experience—from 'Beginners' to 'Veterans' is compiled for each market.

Buyer Diversity

For a sense of the distribution of buyer participation over the auctions in the data set, the number of auctions in which each user placed a bid is tracked. Typically, each identified user participated only in a small percentage of the total number of auctions. Therefore the absolute number, rather than the percentage, is used. Again, four ranges are arbitrarily chosen: [1], [2 – 5], [6 – 10], [>10]. The proportions of buyers falling within these ranges are recorded.

Buyer Experience

This is completely analogous to the measure of seller experience. The feedback ratings of buyers identified in the data set are classified into the same four categories, and the proportions of buyers at each level are compiled for each market.

Matching

Since the intent of auction markets is for competing bidders to arrive at a fair value for the item offered, an auction that ended with a single, winning bid does not make for an interesting or typical case. However, it does signal a unique match in supply and demand.

Dueling

In contrast to single-bid matching, dueling is an indication of keen competition. While the actual dynamics of such bidding patterns, in terms of who is challenging and who is responding, can get quite complex and combinatorial, a tractable criterion can be defined for the occurrence of 'dueling' between two preeminent bidders. Dueling is recognized when the two bidders with the highest number of bids together account for the majority (or some other pre-set level) of the bids, provided the total number of bids exceeds some pre-set threshold (6 is used throughout this study). The last provision is to exclude trivial cases, for example when there are only two bidders with one bid each. Note that the two most active bidders may not actually have encouraged in head-to-head dueling, and neither of them needs to be the eventual winner.

Stashing

In markets for items whose value will increase with scarcity, such as collectibles, certain buyers may set out—to whatever extent—to 'corner' the market. They hope to build up a stock and profit from higher valuations in the future. Or there may be professional dealers with brick-and-mortar sales outlets who have managed to arbitrage actual price differentials between eBay and the conventional marketplace. In either case,

'stashing' activities would be reflected in frequent purchases of the same or similar items within the time frame of the dataset by such buyers. As an indicator of stashing, the highest percentage of the auctions in the dataset won by any single buyer is used.

Sniping

The specific design of online auction will affect the participants' behavior, and hence the topological shape of the market. Three parameters for the eBay online auction system (second price, fixed deadline, and public) lead to an interesting and significant phenomenon known as "sniping" [9]. This is the strategy of placing a first bid at the very end of the auction. In this work, we consider an auction to be won by sniping if the winning bid was placed within one minute from termination, and it was the first and only bid placed by the bidder.

Retailing

On eBay, sellers are provided with an option to offer the item at a fixed price. This is known as Buy-It-Now (BIN). While this option clearly detracts from the true spirit of online auctions, its viability is certainly self-regulating, and its effectiveness affords us another glimpse of the shape of the market.

Proxy

On eBay, a user can submit a maximum bid of what he or she is willing to pay for the item. Then, as necessary, the system will place the appropriate incremental bid for the user against other incoming bids, until the maximum is reached. Auctions won by proxy bids are indicated by winning bids that are not the last ones placed chronologically.

3. A BUYER-SELLER DICHOTOMY FOR ONLINE AUCTION MARKETS

Our topological model is based on the star plot for displaying multivariate data with an arbitrary number of dimensions [1]. Each data point is represented as a star-shaped figure (or glyph) with one ray for each dimension. As the resulting shapes depend on the configuration of the dimensions, we further analyze the observations along the dimensions identified above in an effort to present a visual model (see a generic example in Figure 1) of the shape of online auction markets. This is done by partitioning the dimensions into a buyer-seller dichotomy.

Gross and net activities are significant for inter-market comparisons, and neutral to buyers and sellers. We map gross activities, on some relative scale for all online auction markets, on the upper vertical axis, and use it as the radius of a circle to frame the visual model. Net activities, as percentage of auctions with bids, are mapped on the same vertical axis. Participation, expressed as the average numbers of bids per auctions, is mapped on the lower vertical axis. This way, the size of the circle represents the size of the market relative to

all others, whereas the span between the net activity and participation indicates its vibrancy.

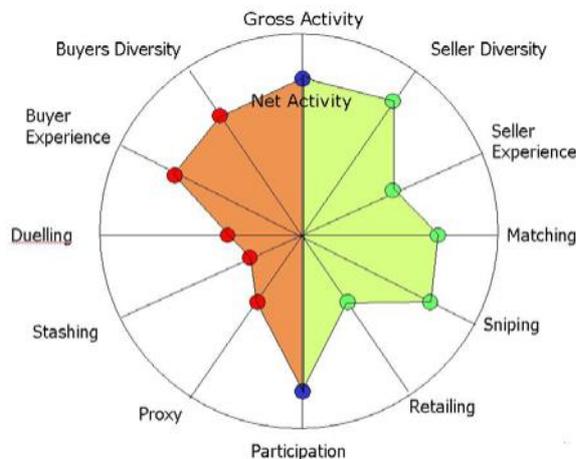


Figure 1. Topology of Online Auction Market

3.1 Buyer dimensions

The seller profiles, in terms of diversity and experience, are buyer dimensions. For diversity, we make the assumption that dominance in market share by relatively few sellers does not provide as much sourcing opportunities for buyers as when there are more different sellers. Analysis of data from over thirty markets led us to propose the use of the proportion of sellers offering 1 per cent or less of the auctions as the metric for diversity. For experience among sellers, the rationale is that an even distribution is favorable to buyers. The skewness of any observed distribution can then be normalized against the even distribution, which is top of the scale. In particular, the maximum of 0.25^4 which equals 0.0039 is normalized to 1.0 so that a distribution of, say, (0.4, 0.3, 0.2, 0.1) will be normalized to 0.61 on this scale. The two seller profile dimensions are mapped radially in the top-right quadrant in Figure 1.

Single-bid matching, though not an indicator of the vitality of a true auction market, does work in the favor of buyers. It meets their demand at opening bid prices, which are likely to be bargains. The success of winning bids by sniping also put buyers at an advantage. As demonstrated in [9], the winning bid prices can be expected to be undervalued. Buy-It-Now options, if acceptable to buyers, necessarily reflect favorable prices, and so can be ruled a buyer dimension. These three buyer dimensions, all expressed as percentage of total auctions, are mapped in the lower right quadrant in Figure 1.

3.2 Seller Dimensions

The buyer profiles, in terms of diversity and experience, are seller dimensions. Analogues to the buyer dimensions we make the assumption that diversity in participation and an even distribution in experience

among buyers are favorable to sellers. For diversity, we use the proportion of buyers participating in a single auction as the metric. For experience, the skewness of the distribution is normalized against the even distribution as described above. These two dimensions are mapped in the top right quadrant in Figure 1.

Dueling, which signals competitive bidding to raise the winning price is obviously to the advantage of sellers. Stashing by buyers with ulterior motives, which are not necessarily rational in the actual market, is also a seller dimension. Finally, proxy bidding can both attract bidders who would otherwise be too busy to monitor an auction, as well as elicit true valuations from potential buyers to help maximize the final price, and hence a seller dimension. These three seller dimensions, all expressed as percentage of total auctions, are mapped in the lower left quadrant in Figure 1.

4. A GLOBAL COMPARATIVE STUDY

The graphical display of the topological model of online auction markets provides a visual aid in describing the 'shape' of a market at any particularly stage of its development. The dimensions (or attributes) deemed advantageous to buyers and sellers are mapped on the right and left sides of a circle, respectively. With proper normalization and harmonization of the scales, the relative areas of the two sides of the star glyph provide a visual cue to market efficiency: a larger right side favors buyers, and vice versa. By themselves, such snapshots encapsulate market characteristics, especially when large numbers of markets are to be studied. As a full-scaled example, a global comparative study is conducted.

4.1 The Countries/Regions

By August 2004, eBay has expanded globally to 23 country/region-specific sites along with its main, US-focused home site. Some of these operate on the same technical platform as the main site. Others, being partnerships or "buy-outs" of local ventures, typically in Latin America and Asia, do not. To date, only sizable sites using the home platform offer recent historical data on completed auctions on an approximately two-week rolling basis. After eliminating sites where activities are not yet high enough to provide any significant data collection, eight are chosen: United States (US), Canada (CA), Australia (AU), United Kingdom (UK), France (FR), Italy (IT), Spain (ES), and Taiwan (TW).

As further evidence of how e-business blurs conventional geo-political distinctions, the scope of the auction markets under any country-specific site is not strictly defined. In general, they offer listings in the local language submitted to the local site, from sellers with items that are located within the specific country. However, it is a seller's option whether to deliver outside of the country or region, and there is no

restriction on where a seller can post a listing. For instance, we might find a German supplier of musical CDs listing offers on, say, the Taiwan site. Ultimately, the relative scope of various markets will be determined by market forces in the form of shipping costs, ease of payment, and language preferences.

4.2 The Markets

Since eBay provides over 27,000 categories of auction items, it may seem trivial to define a market a priori, e.g. digital cameras, and expect to collect data for such a specific item. Actually, the underlying database is keyword driven so that substantial cross-listing can result for items which may only be remotely related. Even with progressively refined drilling to filter extraneous selections, it is impossible, short of item by item scrutiny, to ascertain exact pertinence of the collected data to the prescribed market. For example, even if the search for digital cameras were restricted to a specific brand and model, it would still fail to preclude accessory items such as carrying cases. If a filter is used to block out carrying cases, then it may exclude cameras with carrying cases. With this in mind, the choice of markets for this study is made in part to minimize such fuzziness, as well as to contrast items with common versus subjective values, practical use versus amusement, technology versus craftsmanship. Moreover, they have to be sufficiently active on most of the chosen sites to provide significant data. Even with these criteria, the options are enormous and the decision is arbitrary.

Digital Cameras:

To sharpen the market focus, models with medium pixel capacity is specified. On most sites, these are classified under 3.0–4.0 Megapixels. For ES and UK, 2.1–4.0 is used; 3.1 and over for FR. No subcategory is available for TW.

Diamond Rings:

These are under the category of “Jewelry & Watches” on all sites. While a specific subcategory is available on some sites, a keyword search on “Diamond” may be necessary under “Rings” on others.

CDs Classical:

These are listed under the category of “Music” or “Music and Instruments” and the subcategory of “CD” and genre of “Classical”.

Star Wars Toys and Games:

These are mostly so-called action figures, memorabilia, and collectibles engendered by the Star Wars movies, under the “Toys and Games” category.

4.3 The Data Samples

As we intend to study auction markets at particular points in their development, the time span for the collected data should not be too long. However, to

ensure accurate representation of the actual dynamics of the markets, enough data points should be included. From both logistical and statistical considerations, we chose a target size of 500 auctions over a period of from one to three weeks, in June and July 2004. Note that these are not random samples, but complete records taken over a contiguous interval of time. Not all the actual markets can support this sample size within the given time frame. In those cases, the size is limited to data availability. The sample sizes for this study are summarized in Table 1.

Table 1. Sample size of datasets

Sample Size	US	AU	CA	UK	FR	IT	ES	TW
<i>Digicam</i>	500	444	495	498	100	152	151	426
<i>Diamond Ring</i>	543	504	501	337	98	22	8	10
<i>CDs Classical</i>	501	301	402	500	159	66	-	36
<i>Star Wars</i>	500	587	502	506	602	13	174	20

4.4 The Results

The bidding history for completed auctions on the eBay sites are copied and stored in MS Excel spreadsheets. Macro programs are designed to perform the data mining and compile statistics for the dimensions in the topological model. To render meaning to the visual cue for the buyer-seller dichotomy, the statistics are normalized and harmonized by a transformation that maps (approximately) the first, second, and third quartiles for the entire collection of datasets on each dimension to the unit-free scale of [0.25, 0.5, 0.75], respectively on the unit interval. This way, we can visualize the relative advantages to buyers and sellers without knowledge of the economic details of the transactions.

The results for our datasets are presented by market in Figures 2-5. Since the primary purpose of this work is to illustrate the topological model as a data visualization tool, we will observe, compare, and discuss market characteristics from this perspective, rather than dwelling on the quantitative or statistical analysis. First we note that in terms of net activity (average number of auctions with bids per day), the US site dominates by orders of magnitude. This is indicated in the bar-chart at the center of each of Figures 2-5.

Markets do exhibit their characteristic shapes. As the US site is most developed, the shapes of the markets there provide useful frames of reference. Digicam has an “upright” shape. Diamond Ring is slanted, from top-left to bottom-right. CDs Classical suggests a shamrock leave. Star Wars is more spade like. For a given market, the resemblance across international sites can still be discernable. Significant variations may be attributed to operational factors, regional or cultural differences, but even more likely to market size and stage of development. For example, the Buy-It-Now option tends to be practicable only for large enough

markets. The one remarkable operational factor concerns the TW site. In its promotional period, which covered the data in this study, all listing and commission fees were waived. As a result, the ratio of total listings to those with bids was extremely high. This effect is likely to vanish as soon as regular fees are charged.

In terms of the buyer-seller dichotomy, there is no clear advantage for each side except for Star Wars, which appears to favor sellers in the larger markets. For the others, we can comment on the relative conditions of the markets just by examining the topological plots. Shoppers of digital cameras should find eBay auctions a reasonable source of supply. The only shortcoming based on our model seems to be a dominance of experienced sellers. For hi-tech new products, this is hardly a concern. For sellers in this market, conditions are good as well. The lack of stashing is no surprise for items that are neither collectible nor expected to gain in value. The viability of markets such as Diamond Ring, which involve high value items that are difficult to evaluate online, attests to the potential of e-business. For buyers, the high level of matching translates to opportunities for finding a ring to one's liking at a relatively low starting bid price. With CDs of classical music, it is essentially a retail market as there is scant competitive bidding. Yet the market is viable for specialized and experienced sellers who hope to profit from volume in spite of low margins.

5. CONCLUSION

We presented a global comparative study of four online auction markets in eight countries as a demonstration of a topological model as a data mining and visualization tool, relying only on operational data. A graphical rendition of the model gives shape to such markets, facilitating the visualization of characteristics such as market efficiency, and changes that may favor either buyers or sellers. Conceptually, the star glyph maps market topology into iconography. Research is ongoing to quantify this approach by optimising the dimensional configuration to maximize the resolution of the buyer-seller dichotomy. While the topology is completely independent of market economics, information on the latter can be used in either a knowledge-base or neural net approach to identify shapes of interest. While only casual observations are made for the present purpose of illustration, the

potential exists for more in-depth analysis in the context of plotting e-business strategies in a networked economy, especially when the online auctioning model becomes more widespread in a business-to-business (B2B) environment.

ACKNOWLEDGEMENT

Santosh Bukitgar designed the MS Excel macros for the data mining used in this work. The EMBA class in e-Commerce Projects at UIC in Summer 2004, contributed to the data collection and analysis in this study.

REFERENCES

- [1] Chambers, J., W. Cleveland, B. Kleiner, P. Tukey, *Graphical Methods for Data Analysis*, Belmont, CA: Wadsworth Press, 1983.
- [2] Ho, J. K., "Topological analysis of online auction markets", *International Journal of Electronic Markets*, Vol. 14, No. 3, pp202-213, 2004.
- [3] Hof, R.D., "The eBay economy", *BusinessWeek*, 25 August 2003.
- [4] Hoffman, P., G. Grinstein, "A survey of visualizations for high-dimensional data mining", in U. Fayyad et al (eds.) *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann, pp47-82, 2001.
- [5] Klein, S., "Introduction to electronic auctions", *Electronic Markets* Vol. 7, No. 4, pp3-6, 1997.
- [6] Lucking-Reiley, D., "Auctions on the internet: what's being auctioned, and how?" *The Journal of Industrial Economics*, Vol. 48: pp227-252, 2002.
- [7] O'Toole, K., "The auction architects", *Stanford Business*, Vol. 69, No. 2, 2001.
- [8] Perkins, A. B., M. C. Perkins, *The Internet Bubble*, (Revised Edition), New York: HarperBusiness, 2001.
- [9] Roth, A. E., A. Ockenfels, "Last minute bidding and the rules for ending second-price auctions: theory and evidence from a natural experiment on the Internet", *American Economic Review* Vol. 92, No. 4, pp. 1093-1103, 2002.
- [10] Turban, E. "Auctions and bidding on the Internet: an assessment", *Electronic Markets* Vol. 7, No. 4, pp7-11, 1997.

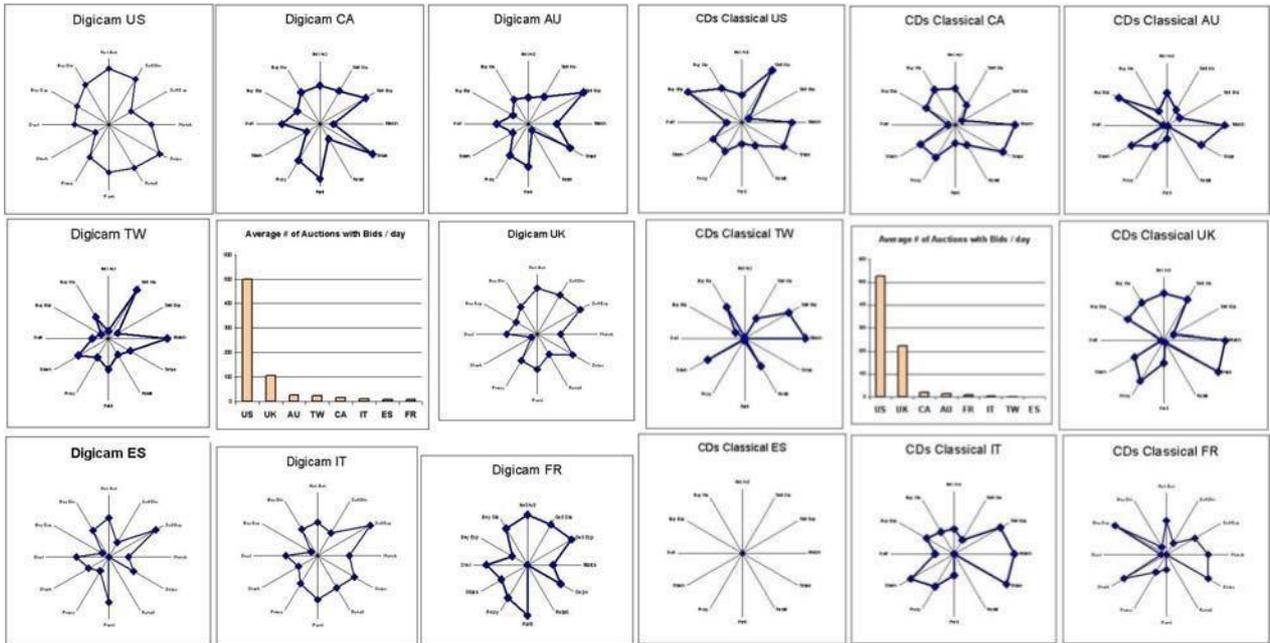


Figure 2. Topologies of Digicam market

Figure 4. Topologies of CDs Classical market

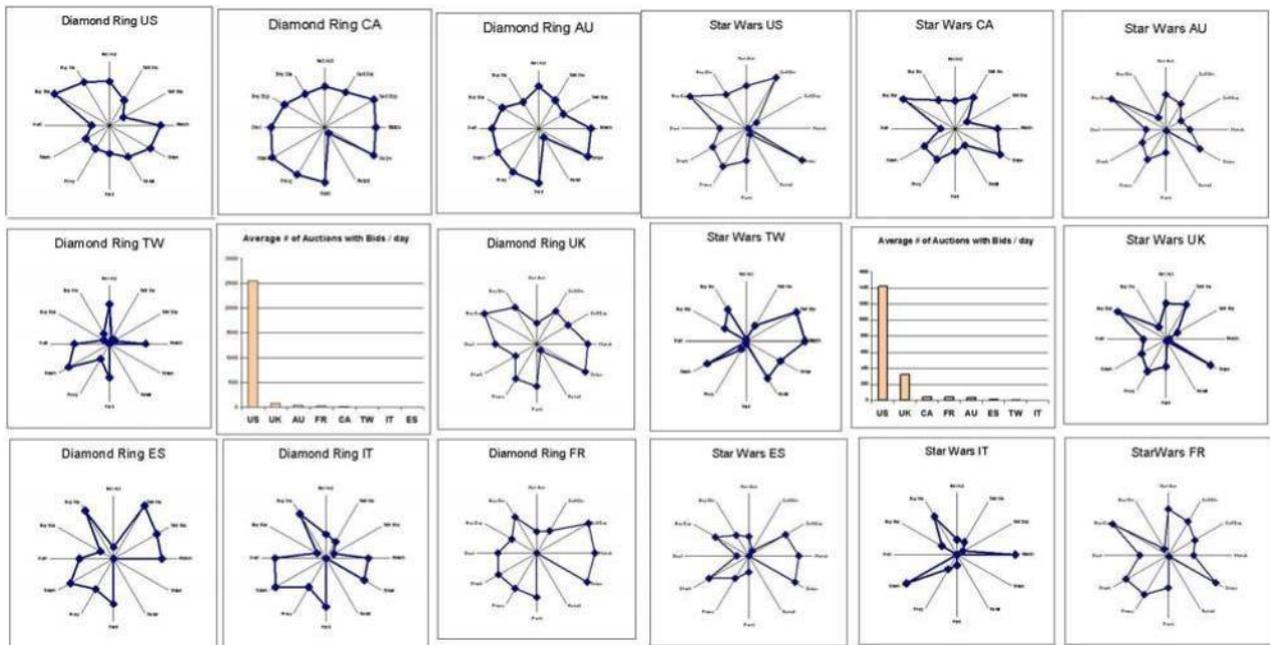


Figure 3. Topologies of Diamond Ring market

Figure 5. Topologies of Star Wars market

A Model of Customer Lifetime Value Consider with Word-of-mouth Marketing Value

Yongmei Liu, Zhenjun Zhao, Yanlong Zhang, Hongjian Lu

Business School, Central South University, Changsha, Hunan 410083, China
ymliu@mail.csu.edu.cn

ABSTRACT

With the rapid development of IT technology and fierce competition of market, the customer relationship management(CRM) has gained its importance in the market. Companies have attached importance to acquiring and retaining the most profitable customers. So calculating customer's value is a significant segment for every effective CRM. Many researches have been performed to calculate customer's value based on customer lifetime value (LTV). But, these calculations can't effectively include the whole customer value, especially for the word-of-mouth marketing value. This paper proposes a new LTV model which considers the customer's past profit contribution, potential value and word-of-mouth marketing value, and gives a more reasonable LTV value in CRM for the company to make a decision.

Keywords: Customer Relationship Management, Customer Lifetime Value, Word-of-Mouth Marketing Value

1. INTRODUCTION

Customer Relation Management (CRM) is a strategy used to comprehend and influence customers' behaviors through effective communication in order to improve the customers' acquisition, retention and loyalty and achieve profitable growth ultimately^[1]. One hot topic in CRM research is how to measure customer value, as the accurately calculating of customer value is the groundwork of identifying and classifying target customers effectively, and it is also the key factors of success of CRM. Customer value is studied through customer equity, customer profitability and lifetime value. In this paper we consider the customer values as the total benefit for corporation during customers' lifetime, namely, LTV—Customer Lifetime Value^[2]. There are already several definitions of LTV, we take Frederick Newell's^[3] here and amend it as:

LTV is, which follows along with the time continuing, the revenues gained from the customers (individual, family or middleman) after the deduction of the total cost of attracting, selling, and servicing customers, and discounted to the present.

It was generally accepted that customer value comprises customers' present value, potential value and loyalty, but we do think the value of world-of-mouth communication is an important component of customer value. Corporations can be benefited as customers recommend productions or services to their friends or family members through world-of-mouth communication way, considering that customers depend more on informal information sources such as acquaintances during decision-making of purchasing^[4-5]. According to a tailed survey to retailing conducted by a Japanese investigating company, 80% saleroom is from present customers, and 60% new customers are from commends of present customers^[3]. Google is a good example: it holds a great deal of users with few commercial advertisements^[6], showing the effectiveness of world-of-mouth communication. Therefore, considers the customer value from four aspects includes current value, potential value, customer loyalty and communication value would be better to indicate the true connotation of LTV.

So, based on the Hyunseok Hwang's conceptual framework of CRM^[7], a CRM conceptual framework based on LTV is shown in Fig.1.

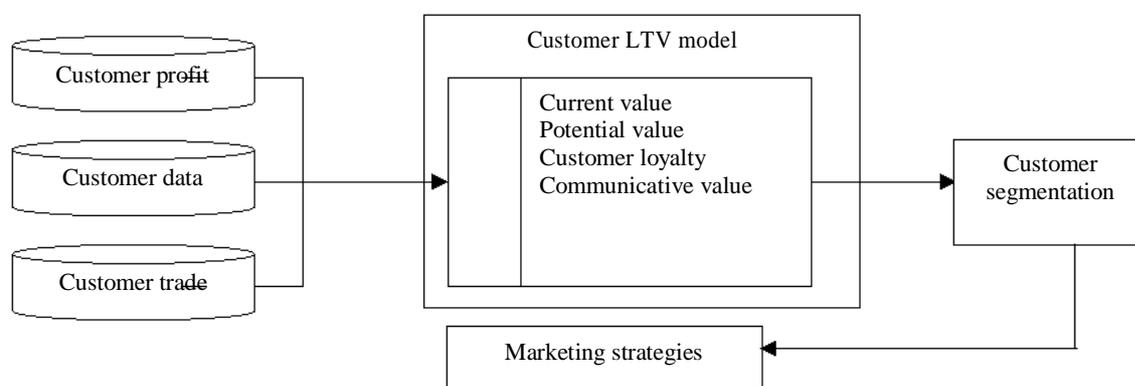


Fig.1. CRM framework based on LTV

Fig. 1 reflects main flow of CRM and the important roles of customer value composing and measure in CRM. Although the measure of LTV is a key factor of CRM, but most corporations haven't got LTV data so far. One important reason of this situation is the complexity of factors, which can influence calculating of LTV. This paper reviewed past LTV models and analyzed factors which influence LTV, proposed a new LTV model considers the customer's past profit contribution, potential value and word-of-mouth marketing. This model can give a more reasonable LTV value in CRM for the company to make a decision, and provide an academic foundation for corporations to manage customer relation better.

2. RELATED RESEARCHES ABOUT LTV

There are a lot of researches on calculating customer value, some researches, such as Huang Yixiao, try to establish an evaluation system by using neural network [8], but most researchers follow Frederick FR.'s net present value (NPV) evaluation system [3]. However most LTV models based on NPV evaluation system exist various limitations, it's difficult for them to reflect the true customer value. The basic LTV model [9] formed based on revenues of loyal customers as follows:

$$LTV = \frac{\ln \left[1 - \frac{1}{(1+r)^n} \right]}{r} \quad (1)$$

Where \ln are revenues that the corporation gain from the loyal customers, r is interest rate, n is the number of years that the customer is loyalty for the corporation. The virtue of the model is its simple calculation, and the disadvantage is that it based on the future revenues' calculation. The measure didn't consider the cost and tend to lead LTV to crossroad.

Considering this situation, Berger and Nasr (1998) have proposed LTV calculation model [7], which reflect the fluctuation of sales and costs.

$$LTV = \sum_{i=0}^n \pi(t) \times \frac{1}{(1+r)^i} \quad (2)$$

Where i is the period of cash flow customer transactions, n is total number of periods of projected life of the customer under consideration, and r is interest rate, $\pi(t)$ is the function of customer profits according to time t . Formulating precise $\pi(t)$ is the most important factor in calculating LTV precisely. LTV models evaluate the long-term value of customers focused on the entire lifetime of customers, however, the long-term value does not fit for the industry having stiff competitions and rapid changes of market environments, the wireless communication industry is good example, which are very sensitive to the external environments and the customer defections. Customer defection is also a critical issue of LTV model because it affects the length of service period and the future profit generation.

Therefore, Hyunseok Hwang suggest a LTV model considering churn rate [7]:

$$LTV_i = \sum_{t_i=0}^{N_i} \pi_p(t_i) (1+r)^{N_i-t_i} + \sum_{t_i=N_i+1}^{N_i+E(i)+1} \frac{\pi_f(t_i) + B(t_i)}{(1+r)^{t_i-N_i}} \quad (3)$$

- t_i : service period index of customer i
- N_i : total service period of customer i
- r : interest rate
- $E(i)$: expected service period of customer i
- $\pi_p(t_i)$: past profit contribution of customer i at period t_i
- $\pi_f(t_i)$: future profit contribution of customer i at period t_i
- $B(t_i)$: potential benefit from customer i at period t_i

This model ignored the importance word-of-mouth marketing value from customers, especially in cybereconomy era, so it is also limitation for calculating the customer's value.

3. A MODEL OF LTV CONSIDERED THE WORD-OF-MOUTH MARKETING VALUE

3.1 A New LTV Model

The word-of-mouth communication has been concerned by people, for the word-of-mouth communication can increase corporations' profits by attracting the new customer for corporations. Suppose the customer i has a little profit for a corporation, but he (she) can attract many more profitable customers for the corporation by word-of-mouth communication. So the customer i should have high customer value. However the models above can't consider this situation, and therefore we propose a model as follows:

$$LTV_i = \sum_{t_i=0}^{N_i} \pi_p(t_i) (1+r)^{N_i-t_i} + \sum_{t_i=0}^{N_i} S_p(t_i) (1+r)^{N_i-t_i} + \sum_{t_i=N_i+1}^{N_i+E(i)+1} \frac{\pi_f(t_i) + B(t_i)}{(1+r)^{t_i-N_i}} + \sum_{t_i=N_i+1}^{N_i+E(i)+1} \frac{S_f(t_i)}{(1+r)^{t_i-N_i}} \quad (4)$$

Compare with Eq.(3), Eq.(4) adds two parts

$$\sum_{t_i=0}^{N_i} S_p(t_i) (1+r)^{N_i-t_i} \quad \text{and} \quad \sum_{t_i=N_i+1}^{N_i+E(i)+1} \frac{S_f(t_i)}{(1+r)^{t_i-N_i}}$$

they denote the net present value(NPV) of customer's past word-of-mouth communication and the future word-of-mouth communication respectively, where $S_p(t_i)$ is past word-of-mouth communication value of customer at period t_i , and $S_f(t_i)$ is future word-of-mouth communication value of customer at period t_i .

3.2 Calculating $S_p(t_i)$ And $S_f(t_i)$

In order to calculate the customer's word-of-mouth marketing value, we divide it into direct and indirect word-of-mouth marketing value. Suppose there is a customer i , then we define direct word-of-mouth marketing value S_0 as a profit contributed by customer i attracting new customer for corporation directly; indirect word-of-mouth marketing value S_1 means a portion profit contributed through attracting new customers for corporation by the customers those who directly attracted by customer i . Then we have:

$$S_p(t_i) = S_{p0}(t_i) + S_{p1}(t_i) \tag{5}$$

$$S_f(t_i) = S_{f0}(t_i) + S_{f1}(t_i) \tag{6}$$

Where $S_{p0}(t_i)$ and $S_{p1}(t_i)$ mean direct and indirect word-of-mouth marketing value of customer i at the past period t_i respectively. $S_{f0}(t_i)$ and $S_{f1}(t_i)$ mean direct and indirect word-of-mouth marketing value of customer i at the future period t_i respectively. We can calculate S_0 and S_1 as follows:

$$S_0(t) = m(t) \cdot c \tag{7}$$

$$S_1(t) = \{E(X_t) - E(X_{t-1})\} \cdot p \cdot \lambda \cdot m(t) \cdot c \tag{8}$$

Where $m(t)$ is the number of customers that directly attracted by customer i at period t , $E(X_t)$ is the expect number of customers from period 0 to period t . And c is the cost that the corporation attracts a new customer by itself. λ is a parameter represents profit sharing ratio of indirect word-of-mouth marketing value (λ can fix on by consulting percentage take from sell profit and relational experience). $p(t)$ is ratio that the new customers attracted by customers in the whole new customer at period t , namely the ratio of attracting new customers by word-of-mouth communication, generally speaking, for different commodities, the word-of-mouth communication have different effects [10]. To calculate $S_1(t)$, we simplified $p(t)$ by p , which is a certain constant. Then $S_1(t)$ can be calculate by

$$S_1(t) = \{E(X_t) - E(X_{t-1})\} \cdot p \cdot m(t) \cdot c \cdot \lambda \tag{9}$$

Word-of-mouth communication has the same characters as virus spread, as a result, word-of-mouth communication marketing has been called virus marketing [11]. So we use the epidemic model [12-13] to calculate $E(X_t)$, and also discrete mathematics [14] model is used to build the word-of-mouth communication model. Now we consider a group of consumers with N people in a certain region, then N means the maximum consumers can be attracted as new customers for a corporation (attracted by either customers or corporation itself). We number the group consumer with a_1, a_2, \dots, a_N , and define set

$A = \{a_1, a_2, \dots, a_N\}$. The word-of-mouth communication among consumer is very complex, we simplified it here and define it as a binary relation on A , i.e. $R_A = \{ \langle a_i, a_j \rangle \mid a_i, a_j \in A, i \neq j, a_i \text{ transmits information to } a_j \}$, we call a_i transmitter and a_j receiver in the binary relation. Then we define set $C = \{a_i \mid a_i \in A, a_i \text{ has been the customer of corporation}\}$. Research is start from the beginning of the production(or service) come into group consumer A , at this time, there is a people a_i among the group consumer who buy the production(or service) and become the first customer of the corporation. Then the condition that another people a_j among the group consumers is attracted by the first people a_i can be described as:

$$\langle a_i, a_j \rangle \in R_A \cap (a_i \in C) \cap (a_j \notin C);$$

Let X_t be the number of customers of the corporation in A at period t , then $\{X_t, t \geq 1\}$ forms a discrete stochastic process. And $E(X_t)$ is the number of the corporation's customer at period t , $(N - E(X_t))$ is the number of potential customers of the corporation. Then during the M times transmissions between period t and period $t+1$, $\frac{M}{N} \times E(X_t)$ is the expected number of people who are not only the transmitter but also the customer of the corporation, $M \cdot \frac{M}{N} \times E(X_t)$ is the expected number of people who are receivers but not belong to set C . According to the hypothesis and stochastic process of transmission, after the series of transmissions, $\frac{M}{N} \times E(X_t) \times \left(1 - \frac{E(X_t)}{N}\right)$ is the expect number of new customer of corporation that attracted by the customer. Then we have:

$$\begin{aligned} & (E(X_{t+1}) - E(X_t)) \times p \\ &= \frac{M}{N} \times E(X_t) \times \left(1 - \frac{E(X_t)}{N}\right) \\ &= \frac{M}{N^2} \times E(X_t) \times (N - E(X_t)) \end{aligned} \tag{10}$$

Let

$$k = \frac{M}{N^2 p} \tag{11}$$

Then

$$E(X_{t+1}) - E(X_t) = k \times E(X_t) \times (N - E(X_t)) \tag{12}$$

So

$$\frac{E(X_{t_i+1}) - E(X_{t_i})}{(t_i + 1) - t_i} = k \times E(X_{t_i}) \times (N - E(X_{t_i})) \quad (13)$$

Regard $E(X_{t_i})$ as a discrete value of a continuous function $f(x)$ at $x = t_i$, according to Eq.(13), the differential equation of $f(x)$ is:

$$\frac{df(x)}{dx} = k \times f(x) \times [N - f(x)] \text{ and } f(0) = 1 \quad (14)$$

Solve the equation, we can get

$$E(X_{t_i}) = \frac{N}{1 + (N - 1) e^{-t_i \times k \times N}} \quad (15)$$

4. CONCLUSION

This paper analyzes factors which influence LTV and their methods of influencing at first, points out that, among all researches on calculating LTV of a customer, no one of them considered about the word-of-mouth marketing value. And then the paper suggests a new LTV model considering the past profit contribution, potential profit, customer loyalty and word-of-mouth marketing value. The model gives a more complete calculation of customer LTV and can be used for customer segmentation especially in the market which word-of-mouth marketing can make good effects.

REFERENCES

- [1] Ronald S. Swift, Accelerating Customer Relationships, translated by Yang Donglong, Beijing, China Economy Press, pp15-16, 2001.
- [2] Zhang Guofang, Jin Guodong, "Survey of Application and Theory Research of CRM", Science & Technology Progress and Policy, Vol.3, pp176-179, 2003.
- [3] Wang Sufen, Tang Bingyong, "Analysis Of Customer Lifetime Value", Journal Of Donghua University (Social Sciences), Vol.2, No.2, pp32-34, 2002.
- [4] Huang Xiaojun, Xu Weiqing, "Reflections On The Word-of-Mouth Communication Research", Journal of Zhejiang University (Social Sciences), Vol.34, No.1, pp125-132, 2004.
- [5] Christian Derbaix, Joelle Vanhamme, "Inducing Word-of-mouth By Eliciting Surprise — A Pilot Investigation", Journal of Economic psychology, Vol.24, pp99-116, 2003.
- [6] Li Zhengquan, "Word-of-mouth Marketing of Google", Foreign Investment In China, Vol.3, pp48-50, 2003.
- [7] Hyunseok Hwang, Taesoo Jung, Euiho Suh, "An LTV Model And Customer Segmentation Based On Customer Value : A Case Study On The Wireless Telecommunication Industry", Expert Systems with Applications, Vol. 26, pp181-188, 2004.
- [8] Huang yi-xiao, Shao pei-ji, Li jing-jing. "The Research About Customer Segmentation Method Based On Customer Value", FORECASTING, Vol.23 No.3, pp31-35, 2004.
- [9] Wang Su-fen, Tang Bing-yong, "The Study Of The Model Of Customer Life-time Value", Journal of Natural Science Heilongjiang University, Vol.19, No.2, pp24-28, 2002.
- [10] Wei Guihua, "The Power of Word-of-mouth", Finance And Accounting Monthly, Vol.06X, pp5-7, 2002.
- [11] Li Guangdou, "Noah's Ark and New Marketing Tools", China Marketing, Vol.7, pp80-82, 2003.
- [12] Lu Li-hua, Luo Peng-fei, Li Xing, Sun Ai-ming, "Summary Of Propagation Model For Network Malicious Mobile Code", Computer Applications, Vol.23, No.6, pp17-22, 2003.
- [13] Pan Jian-ping, Ding Wei, Gu Guan-qun, "A Mathematical Description On The Propagation Of Computer Virus", Journal Of Southeast University, Vol.26, No.6A, pp138-144, 1996.
- [14] George F. Simmons, Introduction To Topology And Modern Analysis, New York, MvGaaw-Hill Book Company, pp3-49, 1963.

An Exploratory Study of the Effects of CRM Practices on CRM Effectiveness and Business Performance

Jashen Chen¹, Russell, K.H. Ching², Eldon Y. Li³, Yiling Liao¹

¹ Department of Business Administration, Yuan Ze University, 10668, Taiwan, China

² Management Information Science Department, California State University, Sacramento
Sacramento, California 95819-6088, USA

³ College of Informatics, Yuan Ze University, Taiwan, China
jchen@saturn.yzu.edu.tw, chingr@csus.edu

ABSTRACT

Continual advances in information technology (IT) have opened new business opportunities in global marketplaces. As a result, many businesses have turned to CRM to gain greater insights into their customers and apply this knowledge toward forging long-term relationships with them. This study examines the relationships of CRM practices (marketing and operational programs) with three antecedent elements (IT investments, absorptive capacity, strategic alignment), CRM effectiveness and firm (business) performance. The results of a survey suggest the following: absorptive capacity and strategic alignment have positive effects on CRM practices, CRM practices affect CRM effectiveness, CRM effectiveness affects firm performance, and CRM effectiveness mediates the effect of CRM practices (marketing programs) on firm performance. Hence, the CRM practices a business adopts will have an impact on its performance.

Keywords: customer relationship management, information technology, absorptive capacity, strategic alignment, CRM effectiveness

1. INTRODUCTION

Continual advances in information technology (IT) have opened the world of electronic commerce (e-Commerce) for businesses to pursue global marketplaces. Predictions for 2005 suggest that worldwide e-commerce will exceed \$5 trillion USD, which will be fueled by over a billion Internet users [13], [17], [25]. In contrast, International Data Corporation estimated the e-commerce market at \$1 trillion for 2003.

Yet, as more businesses enter the Internet-enabled electronic marketplace, competition becomes keener. In seeking means that provide greater competitive advantages, many have turned customer relationship management (CRM), commonly described as an information system to assist the customer retention process or a methodology that extensively employs IT, particularly database and Internet technologies, and enhance relationship marketing (RM) effectiveness and operational practices.

CRM helps a business identify its most valuable customers, and develop, integrate and focus its competencies on forging valuable long-term relationships that deliver superior value to them [20]. It behooves the business to identify and understand these customers to ensure future performance and profitability. However, achieving success also requires investments in technological resources and a clear vision of the business' direction and purpose. Consequently, an understanding of the relationship among the factors that affect and influence CRM practices and performance underlie the achievement of CRM effectiveness and eventually the business'

performance.

This study examines the relationships surrounding CRM practices and its eventual effect on a business' performance. It proposes that three vital antecedent elements, IT investments, organizational absorptive capacity and strategic alignment, shape the CRM practices a business adopts, such as those for its marketing and operational programs. Furthermore, these CRM practices will impact its CRM effectiveness and eventually its performance. Because the three elements reflect the degree to which the business has enriched itself, they represent the building blocks that extend the capabilities of CRM. However, in this successive relationship, CRM effectiveness will mediate the effects of its practices on its business performance. Investments to the practices will affect performance but only if they impact CRM effectiveness. Furthermore, the competitive nature of the industry will moderate the effects of CRM practices on CRM effectiveness. Because businesses respond to market pressures according to the competitive intensity of its industry, practices will vary. Given the interactions of these factors, *how do investments in CRM impact the business' performance?*

2. BACKGROUND

CRM provides a business with an enterprise approach to understanding and influencing customer behavior through meaningful communication. It involves targeting the right customer with the right product or service through the right channel at the right time (Swift, 2001), attracting (acquiring), developing and maintaining successful customer relationships over time

[4], [7], and building customer loyalty [15] through efficient and effective two-way dialogues [19]. As a result, CRM improves a business's ability to acquire and retain customers and build their loyalty, a competitive asset [8].

The insights a business gains through CRM lead to customized products and services that more precisely match its customers' needs and impede competitors from duplicating, imitating or substituting them [18], [24]. With these refined and appealing offerings, a business can expect not only greater satisfaction and continued patronage from its customers, but other benefits in the form of reduced costs and sustained profitability.

Through advanced computing power and high-speed communication networks over the Internet, CRM allows businesses to continually learn about and monitor their customers' needs and expectations, and innovatively respond to them [19]. As a result, businesses can adopt more customer-centric marketing strategies (as opposed to mass-marketing strategies). IT also allows an organization to quickly collect, store, manipulate and disseminate information throughout its operations and with other businesses, such as its supply chain partners, to increase its organizational knowledge and improve its operational efficiencies. IT-based knowledge discovery techniques (i.e., marketing intelligence), such as data mining, online analytical processing (OLAP) and intelligent agents provide the means for building organizational knowledge and further leveraging it for a competitive advantage or profitable gain. Operational benefits involve activities and programs targeting efficiency gains and profitability.

CRM embodies many of the concepts of relationship marketing (RM), a marketing paradigm that focuses on developing close personal relationships, interactions and social exchange between a business and its customers and business partners over time to enhance the its (business') competitive response to continually changing markets [26]. CRM expands upon this concept through its emphasis on information management [18]. Thus, CRM helps focus the business on achieving personal long-term relationships with its customers through information gathered to form a deep understanding of their characteristics, behaviors, needs and expectations, and draw closer to achieving a one-to-one relationship. As a management tool, it directs enterprise resources towards the fulfillment of the organization's goals and objectives.

A business' success with CRM also lies with its execution or practices. CRM practices refer to the actions a business takes to retain its current and attract potential customers through personal touches in fulfilling and satisfying their individual needs. Expanding upon the foundational studies of Applegate et al. [3] and Karimi et al. [12], CRM practices can be

divided into two categories: CRM marketing programs and CRM operational programs.

A business's success with CRM will contribute to its (business) performance. However, this depends on the presence of factors and a sequence of relationships that represent different aspects of CRM. This study examines the interactions tied to achieving success with CRM.

3. RESEARCH MODEL AND HYPOTHESES

Figure 1 illustrates the research model of this study. The model proposes that a business' investments in IT and absorptive capacity, and its commitment to strategic alignment (of IT and the business' strategy) will positively affect its CRM practices (marketing and operational programs), and consequently its CRM effectiveness and (firm) performance. The model also proposes that CRM effectiveness will mediate the effects of CRM practices on firm performance, and industry competition intensity will moderate the effects of CRM practices on CRM effectiveness.

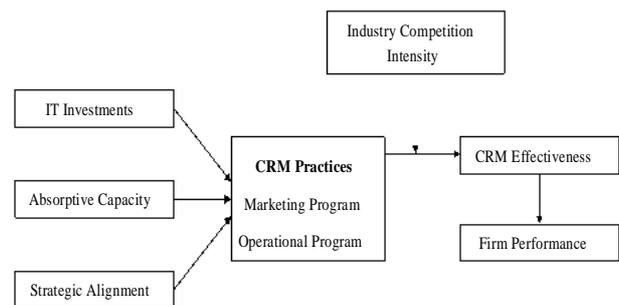


Figure 1 . Research model

Because of CRM's reliance on IT, IT investments determine the extent of CRM's capabilities to support the business' endeavors. The greater the investments the more reasonable it is to expect greater achievements. Thus, IT investments play a major role in shaping CRM practices.

H1: *IT investments have a positive relationship with CRM practices*

H1a: *IT investments have a positive relationship with CRM marketing programs*

H1b: *IT investments have a positive relationship with CRM operational programs*

Organizational absorptive capacity forms the basis for organizational knowledge and the business' ability to innovate. It results from the cumulative learning activities of the organization's individuals and the transfer of knowledge within the organization through a common language [6]. Organizations that have amassed knowledge of their market requirements and

customers' needs are more inclined toward practicing marketing intelligence and information sharing to produce knowledge, and leveraging information and knowledge to satisfy their customers with the goal of gaining their loyalty.

H2: *Absorptive capacity has a positive relationship with CRM practices*

H2a: *Absorptive capacity has a positive relationship with CRM marketing programs*

H2b: *Absorptive capacity has a positive relationship with CRM operational programs*

Because IT enables CRM and extends its capabilities, the strategic alignment between a business' IT and business strategies will impact CRM practices. The alignment reinforces the success of CRM practices. The degree to which strategies align (i.e., strategic alignment) will impact CRM practices, with greater alignment resulting in better practices.

H3: *Strategic alignment has a positive relationship with CRM practices*

H3a: *Strategic alignment has a positive relationship with CRM marketing programs*

H3b: *Strategic alignment has a positive relationship with CRM operational programs*

CRM effectiveness reflects the success of CRM practices. In particular, it examines the outcomes of the business' practices in achieving customer benefits, including social, psychological, economic and customization [9], and business benefits, such as those related to cost reductions, profitability and efficiencies. Thus, greater achievements in CRM practices will be associated with greater heights of CRM effectiveness.

H4: *CRM practices have a positive relationship with CRM effectiveness*

H4a: *CRM marketing programs have a positive relationship with CRM effectiveness*

H4b: *CRM operational programs have a positive relationship with CRM effectiveness*

Prior studies have found a positive link between customer satisfaction and business (firm) performance [1], [2]. Customer loyalty can increase firm performance through the reduction of acquisition and operating costs, and increases in (customer) price tolerance and referrals [11], [16], [22]. Thus, greater levels of CRM effectiveness should have a positive impact on firm performance.

H5: *CRM effectiveness has a positive relationship with firm performance*

Reichheld and Sasser [22] cite a high correlation between customer retention and profitability, such that a five percent increase in customer retention yields improvements in long-term profitability (as stated in terms of net present value) between 20 and 85 percent. Thus, improved CRM practices should have a positive effect on firm performance.

H6: *CRM practices have a positive relationship with firm performance*

H6a: *CRM marketing programs have a positive relationship with firm performance*

H6b: *CRM operational programs have a positive relationship with firm performance*

Businesses will eventually derive benefits from their CRM practices but through CRM effectiveness. Although CRM practices reflect a business' proficiency in leveraging its IT, absorptive capacity and strategic alignment between IT and business strategies, a business will directly benefit from the effectiveness of CRM (i.e., performance attributed to CRM). Given this relationship, CRM effectiveness will mediate the effects of CRM practices on firm performance.

H7: *CRM effectiveness mediates the relationship between CRM practices and firm performance*

H7a: *CRM effectiveness mediates the relationship between CRM marketing programs and firm performance*

H7b: *CRM effectiveness mediates the relationship between CRM operational programs and firm performance*

The role of competition intensity influences a business' responsiveness to customer needs and expectations through product and service offerings [21]. Prior studies have shown the importance of competitive threats to innovation diffusion [5], [14]. Hence, industry competition intensity may influence CRM practices for its effect on CRM effectiveness.

H8: *Industry competition intensity moderates the relationship between CRM practices and CRM effectiveness*

The operational definitions to measure the variables (constructs) were taken from prior studies to test the hypotheses.

4. METHODOLOGY AND ANALYSIS

A survey was conducted on a stratified sample composed of 598 randomly selected companies listed in

the Taiwan Stock Exchange (TSE) (438) and on the over-the-counter (OTC) markets (160). Items were adopted from prior studies and stated on five-point Likert-type scales (1 = strongly disagree, 5 = strong agree). The survey instrument was pre-tested on six experts and managers with extensive CRM practice experiences to elicit their comments and opinions. Subsequent adjustments to improve the questionnaire's clarity were made.

The questionnaires were mailed with a cover letter explaining the purpose of the survey and instructions on its completion, and a return envelope. A total of 161 responses were received for a response rate of 26.9 percent. However, nine questionnaires were returned incomplete and discarded. This reduced the sample size to 153 responses (25.6 percent response rate). A test for sampling bias showed no significant differences among the composition of the sample, the surveyed groups and industry (TSE and OTC). Also, no significant differences were found in a test for non-response bias that compared early and late responses.

A confirmatory factor analysis (CFA) was conducted to assess the discriminant and convergent validity of the items. The values obtained for the measures assess their (measures') validity which is inferred if a measure's factor loadings meet or exceed a criterion value of .40. The results indicate an acceptable level of convergent and discriminant validity.

The maximum likelihood estimation method was used to estimate the structural equation model (SEM) in LISREL 8.3. Figure 2 shows the path loading and coefficient alphas of the constructs. Table 1 summarizes the LISREL estimates and Cronbach alphas of the latent variables.

The analysis of the results generally supports the research model, although not all hypotheses are supported (Table 1). Of the three factors leading to CRM practices, two appear to have positive relationships. The effects of IT investments on CRM marketing and operational programs were not significant (H1 not supported). In contrast, the effects of absorptive capacity and strategic alignment on both CRM practices variables were found significant (H2 and H3 supported). This suggests that organizational factors consistently influence CRM practices, more often than IT investments. Although, IT investments are important to businesses, they may not be a primary consideration to an organization's intended use of CRM. Also, it may be assumed from a practical view point that IT is an integral part of CRM, such that CRM does not exist without IT. This lends further support to CRM as an RM solution and not one of IT.

The data support the positive relationships between CRM practices and CRM effectiveness (H4 supported),

and CRM effectiveness and firm performance (H5 supported). Both marketing and operational programs play vital roles in achieving CRM effectiveness. Marketing programs focus resources on customer sales, service and support, and essential interface elements for ensuring continued patronage. They strive to gain the trust and confidence of the customer. With operational programs, activities center on the logistics of product and service delivery, ensuring their timely and satisfactory delivery. Both programs are important to the longevity of business' relationship with its customers, particularly to maintain customer satisfaction. The support of H5 suggests that maintaining customer satisfaction and loyal (CRM effectiveness) are important aspects of a business that are tendered through CRM. Because CRM practices enhance CRM effectiveness and CRM effectiveness in turn increases a business' performance, improving the practices will have an eventual impact on firm performance.

In the relationship between CRM practices and firm performance, only CRM operational programs have a positive relationship (H6b support); CRM marketing programs are not significant (H6a not supported). Unlike operational programs, the outcomes of marketing programs tend to be intangible and measured in non-monetary terms (i.e., customer satisfaction, loyalty, etc.). Because operational programs objectives and firm performance are gauged monetarily (i.e., costs, profits), both are compatible for measurement under the same terms and sharing a consequential relationship (i.e., one leading to the other).

The results of H4, H5 and H6 (Table 1), and the significant correlations between CRM marketing program and firm performance, and CRM operational program and firm performance suggest that CRM effectiveness mediates the relationship between marketing programs and firm performance (H7a supported) and only partially mediates the relationship between operational programs and firm performance (H7b partially supported). A mediating effect occurs when variations of the independent variables account for variations in the mediator, variations in the mediator account for variations in the dependent variable, and when controlled the independent variables have no effect on the dependent variable. The absence of a direct impact of marketing programs on firm performance (H6a) and the presence of relationships between marketing programs and CRM effectiveness and CRM effectiveness and firm performance indicates the effects have been mediated by CRM effectiveness. However, the presence of positive relationships revealed by H4b, H5 and H6b suggests CRM effectiveness only partial mediates the relationship between operational programs and firm performance.

Although prior studies suggest that competition intensity influences business practices, this study's results do not support a moderating effect of industry

competition intensity on the relationship between CRM practices and CRM effectiveness (H8 not supported). A moderator enhances (or reduces) the effect of the independent variables on the dependent variable. A multiple regression model was developed to test for the moderating effect. Because the interaction of CRM practices and industry competitive intensity is not significant, a moderating effect does not exist (Table 4). Upon further inspection, the data reveal a high mean (4.09) for the measure of competition intensity. This infers that competition intensity does not vary over industries and it becomes indistinguishable. Thus, factors other than competition intensity influence CRM practices.

The maximum likelihood estimation method was used to measure the overall fit of the model (in LISREL 8.3). Hoelter [10] suggests that the data are supportive of the model when the Critical N (CN) exceeds 200. The sample's CN of 519.16 indicates the data appropriately fit the model. The goodness of fitness indicators summarized in Table 5 also indicates a good fit.

Table 1. Summary of test hypotheses

Hypothesis	LISREL Estimates	t-value	Cronbach Alpha	Result
H1a	0.02	0.18	0.957 (IT investments)	Not supported
H1b	0.18	1.78		Not supported
H2a	0.37	4.37	0.922 (Absorptive capacity)	Supported
H2b	0.27	3.58		Supported
H3a	0.41	3.49	0.941 (Strategic alignment)	Supported
H3b	0.65	5.44		Supported
H4a	0.51	7.11	0.787 (Marketing programs)	Supported
H4b	0.46	7.00	0.854 (Operational programs)	Supported
H5	0.27	2.26	0.961 (CRM effectiveness)	Supported
H6a	0.17	1.75	0.927 (Firm performance)	Not supported
H6b	0.22	2.49		Supported
H7a	-	-		Supported
H7b	-			Partially Supported ¹
H8	-	0.272		Not Supported

5. CONCLUSIONS

As more businesses enter the global electronic marketplace and engage in e-Commerce, competition among them intensifies. To better serve their customers and gain their loyal, many have turned to CRM. Understanding the relationships among the elements and components that make up CRM will lead to more effective use of CRM and a greater contribution to profitability. This study has shown that an

organization's investments in absorptive capacity and commitment to aligning its IT and business strategies play largely in forming CRM practices (marketing and operational programs). Significant boosts to CRM practices will lead to greater CRM effectiveness, which in turn contributes to profitability. Thus, businesses entering e-Commerce and investing in CRM should understand the organizational commitments and activities surrounding a CRM implementation. Although IT enables many aspects of e-Commerce and integrates activities across organizational boundaries, its role in CRM other than an enabler and integrator may not be assumed as an essential part of CRM. Thus, achieving CRM effectiveness lies in how the business leverages and exploits its knowledge.

ACKNOWLEDGEMENTS

This study was supported in part by the National Science Council of Taiwan, China.

REFERENCES

- [1] Anderson, E.W., Fornell, C. and Lehmann, D.R. "Customer satisfaction, market share, and profitability: findings from Sweden," *Journal of Marketing*, 58(July), 53-66, 1994.
- [2] Anderson, E.W., Fornell, C. and Rust, R.T. "Customer satisfaction, productivity, and profitability: differences between goods and services," *Management Science*, 16(2), 129-145, 1997.
- [3] Applegate, L.M., McFarlan, F.W., and McKenney, J.L. *Corporate Information Systems Management: The Issue Facing Senior Executives*, 4th ed. Chicago, IL: Irwin, 1996.
- [4] Berry, L.L., and Parasuraman, A.. *Marketing Service*. New York, NY: Free Press, 1991.
- [5] Birkinshaw, J., Morrison, A. and Hulland, J. "Structural and competitive determinants of a global integration strategy," *Strategic Management Journal*, 16, 637-655, 1995.
- [6] Cohen, W.M., and Levinthal D. "Absorptive capacity: a new perspective of learning and innovation," *Administrative Science Quarterly*, 35, 128-152, 1990.
- [7] Day, G.S. "Managing Marketing Relationships," *Journal of the Academy of Marketing Science*, 28, 24-31, 2000.
- [8] Dekimpe, M.G., Steenkamp, J-B E.M., Mellens, M. and Abeele, P.V. "Decline and variability in brand loyalty," *International Journal of Research in Marketing*, 14, 5, 405-420. 1997.
- [9] Gwinner, K.P., Gremler, D.D., and Bitner, M.J. "Relational benefits in services industries: the customer's perspective," *Journal of the Academy of Marketing Science*, 26, 101-114, 1998.
- [10] Hoelter, J.W. "The analysis of covariance structures: goodness-of-fit indices," *Sociological Methods and Research*, 11, 325-344, 1983.

- [11] Kalwani, M.U. and Narayandas, N. "Long term manufacturer—supplier relationships: do they pay off for supplier firms?" *Journal of Marketing*, 59(January), 1-16, 1995.
- [12] Karimi, J., Somers, T.M., and Gupta, Y.P. "Impact of information technology management practices on customer service," *Journal of Management Information Systems*, 17, 125-158, 2001.
- [13] Killen & Associates. "e-Commerce B2B and B2C, essential document-based applications and markets study," www.killen.com, 2001.
- [28] Kobrin, S.J. "An empirical analysis of the determinants of global integration," *Strategic Management Journal*, 12 (Summer), 17-32, 1991.
- [15] Kohli, R., Piontek, F., Elington, T., VanOsdol, T., Shepard, M., and Brazel, G. "Managing customer relationships through e-business decision support applications: a case of hospital-physician collaboration," *Decision Support Systems*, 32, 171-187, 2001.
- [16] Loveman, G.W. "Employee satisfaction, customer loyalty, and financial performance: an empirical examination of the service profit chain in retail banking," *Journal of Service Research*, 1(1), 18-31, 1998.
- [17] Mullaney, T.J., Green, H., Arndt, M., Hoff, R.D., & Himelstein, L. "The e-biz surprise," *Business Week*, (May 12), 60-68, 2003.
- [18] Peppard, J. "Customer relationship management (CRM) in financial services," *European Management Journal*, 18, 312-327, 2000.
- [19] Peppers, D., Rogers, M. and Dorf, R. "Is your company ready for one-to-one marketing?" *Harvard Business Review*, Jan-Feb, pp. 151-160, 1999.
- [20] Plakoyiannaki, E., and Tzokas, N. "Customer relationship management: a capabilities portfolio perspective," *Journal of Database Marketing*, 9(3), 228-237, 2002.
- [21] Porter, M.E. "The competitive advantage of nations," *Harvard Business Review*, March-April, 73-93, 1990.
- [22] Reichheld, F.F., and Sasser, W.E., Jr. "Zero defections: quality comes to services," *Harvard Business Review*, September-October, 105-111, 1990.
- [23] Swift, R. S. *Accelerating Customer Relationships*. Upper Saddle River, NJ: Prentice-Hall, 2001.
- [24] Winer, R.S. "A framework for customer relationship management," *California Management Review*, 43(4), 89-105, 2001.
- [25] ZDNet. "Expect one billion net users by 2005," ZDNet, znet.com.com/2102-11_529900.html, 2001.
- [26] Zineldin, M. "Beyond relationship marketing: technologicalship marketing," *Marketing Intelligence and Planning*, 18, 9-23, 2000.

An Intelligent Model for Stock Investment with Buffett Strategy, Classifier System, Neural Network and Linear Programming

Anping Chen¹, Weili Lin², Yenchu Chen³

¹ Chair, Institute of Information Management, National Chiao-Tung University,
Hsin-Chu, 300, Taiwan, China

^{2,3} PhD Student, Institute of Information Management, National Chiao-Tung University,
Hsin-Chu, 300, Taiwan, China

¹apc @apex.com.tw, ²weililin.iim92 @nctu.edu.tw, ³irene@iim.nctu.edu.tw

ABSTRACT

“The Intelligent Model for Stock Investment with Buffett Strategy, Classifier System, Neural Network and Linear Programming” was studied for developing an intelligent model which can learn more knowledge regarding to stock investment with artificial intelligence technology. Classifier system, neural network, fundamental financial investment factors and linear programming are the fundamental components for the research. Knowledge transformation and genetic evolution capability was discussed in the article, too. Furthermore, the investment strategy developed by Warren E. Buffett[17], the great financial investment master, was the major knowledge which was practiced in the article.

For realizing more detail about learning system, a lot of topics regarding to artificial intelligence were discussed in advanced, including “A Market-Based Rule Learning System” [1], “Dynamic Trading Strategy Learning Model using Learning Classifier System” [2], “Nonlinear Index Prediction” [3], “Financial Decision Support with Hybrid Genetic and Neural Based Modeling Tool” [4] and “Fuzzy Interval methods in Investment risk Appraisal” [5].

According to the study mentioned above, the ideas to give intelligent model, especially with genetic algorithm, bring the direction for the advanced financial investment strategy and operation. Therefore, it was why a novel intelligent model with Buffett strategy, classifier system, neural network and linear programming proposed in the article.

Keywords: Classifier System, Security Investment, Buffet, Neural Network, Linear Programming

1. INTRODUCTION

In the world of financial investment, everybody is looking forward to raising the investment visibility and get profit through transparent financial investment model. However, people are confused by too many kinds of merchandize, and suffered by the high complexity with so many connections among all investors from different security and trading market. It is possible but really difficult to award the profit in the mazy trading environment.

Trading in stock market is the most popular financial investment. The shaky price in stock market brought the possible super profit. The entry barrier to involve trading in stock market is low, too. The market price and related financial information is easy to take. It is the reason why the population of involving stock market is very high in anywhere.

For providing more intelligent technology in financial investment prediction, only the research in classifier system, neural network, linear programming, and fundamental intelligent financial investment operation, but also the investment criteria developed by Warren E. Buffett [6][17][18] was adopted for the improving the intelligence of financial investment operation.

This was why the research including linear

programming, neural network, classifier system, fundamental financial investment factors and Buffett investment criteria were taken in the article. The intelligent model can be constructed by integrating these methodologies. Through the developed intelligent model and related information, the most effective rules can be retrieved for supporting financial investment decision making. The fundamental factors are wisdom-base mining operations in the intelligent model. Therefore, Buffett strategy was taken for representing the knowledge of mid-term / long-term financial investment. Furthermore, in the article the knowledge will not be limited on the idea from Buffett Strategy, but also including any financial knowledge if they will be useful for predicting the stock price in the changeable world.

2. RELATED WORK

2.1 Linear Programming

The Linear Programming (LP) [7] is the model which represents the objective system functions and related behavior. The basic definition of LP can be described as below:

$$\begin{aligned} \text{Object function : } & \text{minimize } cx \dots(1) \\ \text{Subject to } & Ax = b \dots\dots\dots(2) \\ & x \geq 0 \dots\dots\dots(3) \end{aligned}$$

Where x is the multi-dimensional variable vector associated with coefficient matrix A , c , b are coefficient vectors, the formula (2) and (3) are constraint conditions. The formula (1) is the objective function in the linear programming.

Normally, for a specific objective function, there are quite a lot dimensions and variables. To simplify the complexity of given problem, the linear programming model, normally, was constructed with limited factors. So, it is necessary to construct a dynamic intelligent model for intensifying financial investment prediction.

2.2 Neural Network

As what we knew, all the messages what human can feel are dispatched by network system with neuron and delivered to brain for further decision making. Neural network [8] which was defined in information technology simulates the operation, transmits the messages from sender through given activation function in each “cell” to receiver.

As the definition of neural network mentioned above, any cell in neural network normally gives different activation function within it. A sample of a cell in neural network can be depicted as the figure as below:

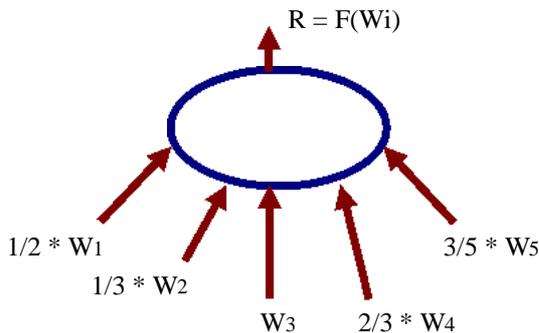


Figure 2 : A Sample of A Cell in Neural Network

where, for each input $W_i = 0$ or 1

$$\begin{aligned}
 F(W_i) &= 2/3, & \text{if } 1/2 W_1 + 1/3 W_2 + W_3 + 2/3 W_4 + 3/5 W_5 \geq 1.5 \\
 &= 0, & \text{if } 1/2 W_1 + 1/3 W_2 + W_3 + 2/3 W_4 + 3/5 W_5 < 1.5 \dots\dots(4)
 \end{aligned}$$

The value R is the return of activation function in the given cell. It got the 5 input sensors and forward to next stage (may be the other cell). The value 1.5 for function $F(W_i)$ is the “Threshold Value” for the cell’s activation function. Each input which is associated with a “weight” can be defined as a factor for decision making in R . Each “weight” for corresponded “input” is normally independent to the other “weight” for the cell in neural network.

2.3 Classifier System

Classifier System [9] is the major research model for evolutionary computation in artificial intelligence. It is an intelligent learning mechanism with rules of expertise, messaging between rules and parallel process. Within the organization of classifier system, for intensifying the message delivery between rules, the Bucket Brigade Algorithm [10] was taken for solving Credit Assignment [11] in system. In addition, for self-creating new rules, classifier system was improved as a new expert system which can be the rules set with genetic algorithm. Each rule adopted in the system is treated as a *classifier*. Therefore, the operational mechanism for the whole classifier system is constructed on the evolutionary concept with information transmission and credit exchange among all classifiers. The system structure of whole classifier system can be depicted as figure 3 :

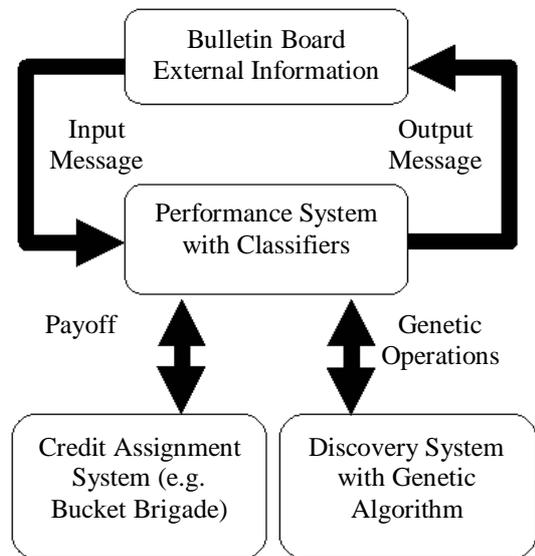


Figure 3 : The System Structure of Classifier System

In the classifier system, the Performance System Module is located on kernel [12]. It is also the system module which connects to message outside. It is looked like an expert system but not limited on specific knowledge domain. Although it is rule-based system module, however, it still needs the additional system which can help to process the outside messages via information transmission, high degreed standardized action and parallel processing capabilities, and can tell the advantages from each classifier in the system.

The 2nd major system in classifier system is Credit Assignment System which can tell the advantages from different classifier, and evolve them via tuning credit. Normally the Bucket Brigade Algorithm was taken for processing the functionality. According to the definition of credit assignment, the Bucket Brigade Algorithm is designed for telling the right rules under a rule-based classifier system. Also by the Bucket Brigade Algorithm, better rules will be rewarded with more credits and worse rules will be punished. Each classifier in the credit assignment system indicates its own credit level.

Through the Bucket Brigade Algorithm the credit level for each classifier will be tuned by corresponded contribution in outside message handling. Also, the classifier can bid and obtain the right to handle outside message if it got highest credit level among all competitive classifiers.

The 3rd major system in classifier system is Rule Discovery System[13]. For the evolution after the competition among all classifiers, the portion of useful classifier is probably the minority in the whole classifier rule base. With the minority portion of classifier, it cannot be the answer to find the best set of classifiers so that the new classifiers (rules) should be generated via certain experience and evolutionary criteria. That is, the new classifiers are expected to generate either from existed excellent classifiers, or changing existed classifiers.

2.4 The Factors of Security Financial Investment

Security financial investment must be defined in term of market behavior. The market behavior, from fundamental viewpoint, is the supply – demand operations in term of economic activities [14]. The pricing is determined by the supply – demand operations, too.

The impact to price from supply – demand operations cannot be briefly explained with marketing trading. For buyers in stock market, the investment can be defined as two portions: stocks on hand and stakes for buying. Of course, for sellers in stock market, the investment can be defined with opposite way: stakes on hand and stocks for selling. Therefore, the timing for tuning stocks and stakes is taken as the strategy of investment.

In addition, for considering the balance of supply – demand relation made by quantity and price in stock market, the factors of analysis including technique, fundamentality, industry development, macro economics, information and stake, are the keys [15]. The motivation for investment including venture, risk prevention and profit locking, drives the balance of trading market in terms of operation model and momentum [16]. These factors above were discussed in the article.

2.5 Buffett’s Criteria in Financial Investment

The philosophy of Buffett’s financial investment is: “Search and invest the healthy enterprise entity if its stock price is much underestimated. Keep the stocks as long-term then the profit generated via compound interest will return.” Therefore, the enterprise that Buffett preferred to invest got the characteristics listed as below [17]:

1. The enterprise gets the position of monopolization. The population of products is good. Business is

reliable and can be developed as well in the future.

2. The enterprise got reliable financial situation. The management performance is excellent. The capital investment is little and cash flow is sufficient.
3. If the enterprise is found by the conditions above, then evaluate the related internal value thereafter. If the current stock price of the enterprise is much less than the value after evaluation (normally the current stock price is 30% less than internal value), then the enterprise is deserved to invest.

When Buffett found the enterprise which met the conditions mentioned above, then he usually ignored the oscillation of stock price, funded quite a lot stakes and bought significant portion of stocks, as the way to purchase the enterprise. Buffett kept these stocks with long-term way. The timing he sold the stock was determined with the following conditions:

1. The current stock price is more than the internal value of the enterprise after evaluation.
2. The growth of aimed enterprise becomes slow.
3. More stakes are needed for better target enterprise.

3. THE OBJECTIVE FUNCTION OF THE INTELLIGENT MODEL

The operation of this intelligent model with classifier-based neural network and Buffett strategy can be explained by Figure 4:

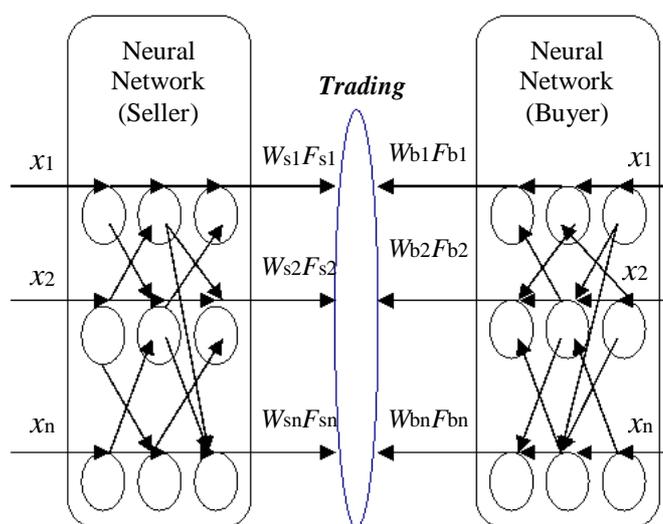


Figure 4 : The Intelligent Model with Classifier-based Neural Network and Buffett Strategy

In Figure 4, $W_{si}F_{si}$ and $W_{bi}F_{bi}$ stand for the investment decision making information for the sellers and the buyers. Take the sellers as an example, F_{si} represents the i^{th} classifier of the seller party in the neural network,

and W_{si} symbolizes the weighted function of the classifier F_{si} . Formula (7) and (8) illustrate the definitions of these two function variables for the buyer party in the investment model, and the objective function is illustrated in formula (5) and (6).

Object function : $minimize |P_r(t) - P(t)| \dots\dots\dots(5)$

Subject to:

$$P(t) = P(t-1) + \sum_{i=1}^M F_{bi}(t-1) - W_{bi}(t) \dots\dots\dots(6)$$

where

$$F_{bi}(k) = PE_{stock_i} \times EPS_{stock_i} \times \sum_{k=t-10}^{t-1} PBT \times \sum_{k=t-12}^{t-1} GRP \dots\dots\dots(7)$$

$W_{bi}(t-1)$ = the proportional weight in all classifiers of stock i in time $t-1$ (8)

$P_r(t)$ denotes the real stock price in time t . $P(t)$ is the expected price after investment. $P(t-1)$ represents the present stock price before investment. The neural function in classifier system is shown in formula (6) where M is the number of stocks that $F_{bi}(t)$ and $W_{bi}(t)$ are adopted. We can observe the behavior of $F_{bi}(t)$ by formula (7). $W_{bi}(t)$ represents the weighted function in time t , as shown in formula (8). The object of this model, as stated in the objective function shown in formula (5), is to minimize the difference between $P(t)$, the expected stock price, and $P_r(t)$, the actual stock price in time t .

Obviously these functions $F_{bi}(t)$, $F_{si}(t)$, $W_{bi}(t)$, and $W_{si}(t)$ which determine the value of $P(t)$, is executed in the neural network. Say for the buyer, $W_{b1}(t)$ is the weighted function of $F_{b1}(t)$, $W_{b2}(t)$ is the weighted function of $F_{b2}(t)$, $W_{b3}(t)$ is the weighted function of $F_{b3}(t)$... and so on. $F_{b1}(t)$, $F_{b2}(t)$, and $F_{b3}(t)$ themselves are analyzable functions, therefore there exists respective parameters for each $F_{bi}(t)$. This kind of nested neural network structure benefits the linear programming model well.

Defining $F_{bi}(t)$ and $W_{bi}(t)$ is the key issue to construct the most effective linear programming model within neural network framework. In this article, we use the "Divide and Conquer" to discover the best factor and choose the weights. The factors in the neural network framework, resulting in the linear programming model, build up a decision support system for the portfolio investment.

However, with the characteristics of dynamic linear programming, the analysis of neural network can be tuned. On the other hand, the static linear programming actually is far away from the real world. The stock investment cannot rely on such kind of model only because it will be distortive when the factors and stock environment start to change.

Certainly, the functions including $F_{bi}(t)$, $F_{si}(t)$, $W_{bi}(t)$ and $W_{si}(t)$ can be the other dynamic sub-objective functions. The precision of the related linear programming is determined by design scope, cost and the degree of detail. In the article, the way to improve the precision of objective function will be built on the realization on definition of dynamic parameters.

4. THE PRACTICE OF INTELLIGENT MODEL OPERATION

There were 6 samples taken for the coming experiment. However, it did not indicate the all criteria of Buffett knowledge. Also, the rules depicted from the given 6 samples do not show up the know how of the whole research of security investment. The intelligent model for security investment with classifier system, neural network and linear programming is designed for inputting more rules or knowledge, such as Buffett strategy and investment criteria, can be input and transformed as the classifiers and be evolved in the intelligent model.

According to related reference regarding to the research of Buffett investment strategies [17][18], the rule of 6 assumed sample classifiers and corresponded format can be defined as below:

1. Classifier_001 : To retrieve stock if current price got 30% difference less than internal value of enterprise.

Condition of Classifier	Action	State
# 0 0 0	0 1	0 1 0 1

2. Classifier_002 : To retrieve stock if current price got 40% difference less than internal value of enterprise.

Condition of Classifier	Action	State
# 0 0 1	0 1	0 1 0 1

3. Classifier_003 : To retrieve stock if current price got 50% difference less than internal value of enterprise.

Condition of Classifier	Action	State
# 0 1 0	0 1	0 1 0 1

4. Classifier_004 : To retrieve stock if current price got 60% difference less than internal value of enterprise.

Condition of Classifier	Action	State
# 0 1 1	0 1	0 1 0 1

5. Classifier_005 : To retrieve stock if current price got 70% difference less than internal value of enterprise.

Condition of Classifier			Action			State			
#	1	0	0	0	1	0	1	0	1

6. Classifier_006 : To retrieve stock if current price got 80% difference less than internal value of enterprise.

Condition of Classifier			Action			State			
#	1	0	1	0	1	0	1	0	1

For the 6 sample classifiers, the input message, output message, classifier function for telling the difference between current stock price and internal value of enterprise, and weighted function of bidding value can be depicted as below:

- ϕ Input Message : including stock ID, PE, EPS, the profit rate before tax for past 10 quarters, revenue growth rate for past 12 months, the current price of the stock.
- ϕ Output Message : same as input message, including stock ID, PE, EPS, the profit rate before tax for past 10 quarters, revenue growth rate for past 12 months, the current price of the stock.
- ϕ Classifier Function to Tell the Difference : ((PE) X (EPS) X (the profit rate before tax for past 10 quarters) X (revenue growth rate for past 12 months)) X (Condition)% >= (current price of stock).
- ϕ The weighted function of bidding value : 10% of the proportional weight in all classifiers.

The factors to evaluate stock value can be depicted with PBT (The Profit before tax), S.P (Current Stock Price) and margin is described as Table 1 as below :

Table 1 : The Factors To Evaluate of Stock Value

EPS	PE	12M GRR	PBT	Value	<Date> S.P	Margin
2.40	20.00	3.04%	-0.24%	49.34	20.5	140.6%

According to the message format defined above, the stock value which is evaluated by classifier can be depicted as formula as below:

$$Stock\ Value = EPS \times PE\ (in\ Industry) \times (12M\ GRR) \times PBT \dots\dots(9)$$

In the Table 1, 12M GRR (Growth Rate in term of Revenue) indicates the growth rate in term of monthly revenue in past 12 months. PBT indicates the profit rate before tax in past 10 quarters. Certainly, the regulation in the classifier is not necessary calculated by mean value but by the other buffett investment philosophy. For easily describing how to transform buffett

investment strategy as a classifier, the assumed factors and regulation mentioned above are taken for designed as “Buffett Classifiers” for practice.

Furthermore, why the both “12M GRR” and “PBT” are taken for designing a type of “Buffett Classifiers”?

1. According to the research of Buffett investment strategy, for preventing the error caused from stock price oscillation, the time period which is taken for information collection must be longer. The time period with 12 months or 10 quarters is a proper and reasonable default value to converge the performance of business operation and stock price.
2. The internal value of stock price for given enterprise, EPS and PE are the fundamental factors. However, the value of enterprise should be assessed according to operation performance and capability to make profit. Therefore, the factors “12M GRR” and “PBT” are taken as the parameter to evaluate the stock price. Note that the Buffett’s attitude of stock investment was similar to “purchase the enterprise” rather than the gamble way.

So, the minimization of difference between **Pr(t)** and **P(t)** is objective function in the article. This is the design target for the proposed intelligent model. The mathematic characteristics in the intelligent model are also the research objectives in the article.

Note that the criteria within the assumed classifiers are not indicated high precision in stock price prediction. Probably some stronger classifiers have shown right assessment via giving internal value for the related enterprise. It must have some other weaker classifiers that never give right prediction. The point in the article is “the environment” for rules evolution from the structure and mechanism in the proposed intelligent model. The stronger classifiers that provide right answer in stock price prediction can be evolved through more investment strategies and learning via the genetic evolutionary process.

5. CONCLUSION

Through the research and practice in the article, the conclusion and additional issues were listed as below :

1. The intelligent model for stock investment with Buffett strategy, classifier system, neural network and linear programming got a lot of advantages in system learning and intelligence evolution. Any kind of valuable knowledge, even wisdom, such as Buffett stragey in stock investment, can be transformed and deliver into the intelligent model.
2. Through the intelligent model for stock investment, especially follow the Buffett criteria, the performance to make profit will not be carried in

very short time period. The basic learning cycle is stilled required.

3. The intelligent model is constructed with classifier system, neural network, fundamental financial factors and Buffett investment strategy. Obviously it integrated the research fields in artificial intelligence and financial investment. Through the study, except the useful rules can be adopted and evolved in the intelligent model, the rules with high performance of financial investment can be retrieved from the system operation.
4. All the samples and assumed classifiers are the basic objects for practice in the intelligent model. With longer practice, around more than half or 1 year, the precision of the intelligent model and stronger classifiers would be found.
5. The oscillation rate in linear credit function can be the advanced issue for research in the intelligent model. Also, the risk problem of the credit distribution among all classifier is deserved to discuss. The information in tendency or habitual domain for each classifier can be reveal from the credit assignment and genetic evolution. It would get the potential risk if no superior advisable intelligent model manages the operational intelligent model.
6. The variance beyond the system prediction should be considered. Therefore, the knowledge for high degree of stock price oscillation should be taken in linear programming.

REFERENCES

- [1] Paper: Qing Qing Zhou (GuanDong Data Communication Bureau, China Telecom), Martin Purvis (University of Otago Dunedin, New Zealand), "A Market-based rule learning system.", ----- ACM International Conference Proceeding.
- [2] Paper: Pen-Yang Liao, Jiah-Shing Chen, "Dynamic Trading Strategy Learning Model Using Learning Classifier System", Evolutionary Computation, 2001, Proceeding of the 2001 Congress on, Volume: 2, 27-30 May, 2001, P783-789.
- [3] Paper: Zemke Stefan, "Nonlinear Index Prediction", Physica A, Volume: 269, July 1st, 1999. P177-183.
- [4] Paper: Kumar Ned, Krovi Ravindra, Rajagopalan Balaji, "Financial Decision Support with Hybrid Genetic and Neural Based Modeling Tools", European Journal of Operational Research, Volume 103, December 1st, 1997, P339-349.
- [5] Paper: Sergrieva Antoaneta, Hunter John, "Fuzzy Interval Methods in Investment Risk Appraisal", Fuzzy Set and Systems, Volume: 142, March 16th, 2004, P443-466.
- [6] Robert P. Miles, (2001), "*101 Reasons to Own the World's Greatest Investment: Warren Buffett's Berkshire Hathaway*", ISBN: 047141123X.
- [7] Text Book : G. Hadley (1997), "Linear Programming", Industrial Management at the Massachusetts Institute of Technology, P.4.
- [8] Text Book (Chinese Version) : M.C. Su, H. D Chang (1997), "Neural Network, Fuzzy System and Genetic Algorithm", CHWA Book Store.
- [9] J. H. Holland, K. J. Holyoak, R. E. Nisbett, and P. R. Thargard (1986), "Induction: Processes of Inference, Learning, and Discovery", MIT Press.
- [10] Thesis : C. F. Chen (2000) , A Thesis for Master Degree, "A Dynamically Adaptive Routing Method Based on Classifier System", Institute of Computer and Information Science, National Chiao-Tung University, P.18 ~ P.19.
- [11] Thesis : C. F. Chen (2000) , A Thesis for Master Degree, "A Dynamically Adaptive Routing Method Based on Classifier System", Institute of Computer and Information Science, National Chiao-Tung University, P.12 ~ P.13.
- [12] Thesis : Dayle Major (2002), An Introduction to Learning Classifier Systems , <http://web.umn.edu/~tauritzd/courses/cs447/project/Majors.doc>, P.2.
- [13] Thesis : C. F. Chen (2000) , A Thesis for Master Degree, "A Dynamically Adaptive Routing Method Based on Classifier System", Institute of Computer and Information Science, National Chiao-Tung University, P.13 ~ P.14.
- [14] Text Book : C.H. Chang, J.D. Hsu, Y.C. Liu, T.M. Wu (1991), "Economics – Theory and Practice", Version 2, Department of Economics, National Taiwan University, P.20 ~ P.44,
- [15] Text Book : A.P. Chen (2002), "Financial Information Management and Investment Decision", APEX Information Corp., P.29 ~ P.30.
- [16] Text Book : A.P. Chen (2002), "Financial Information Management and Investment Decision", APEX Information Corp., P.17.
- [17] Wen-Shiu Lin, Chien-Pei Yang, (May, 2003), "Application of Integral Value-investing Strategy with Genetic Algorithms", Tamsui Oxford Journal of Management Science June 2003, Vol 19, No. 1, (19-50), Aletheia University.
- [18] Wang Chiueh-cheng, 2002. Money Making of New Investment Bible Baphit, Commercial Newsweek issue 769, August.

Analysis of Costs, Benefits and ROI of CRM Implementation*

Lun Hou¹, Xiaowo Tang²

¹ School of Management, UESTC, Chengdu 610054, China

² Education Bureau of Sichuan Province, Chengdu 610000, China

ABSTRACT

In the recent years, customer relationship management (CRM) has being depicted by the software vendors as a cure-alls for the enterprises which wish to acquire, maintain efficiently customer share and at last to gain more profit in globalized market with fierce competition. However, the currently popular CRM marketing practice often produces disappointing outcomes. The more popular this marketing practice gets, the more people who realize that the current CRM practice hardly manages customer relationship. Through analyzing the costs-benefits, ROI and TCO of CRM implementation, this article tries to find out a correct way to accomplish successfully CRM implementation, discuss the success factor and failure barrier of CRM projects for successful complementation of CRM.

Key words: Customer Relationship Management (CRM), CRM implementation, Cost-Benefit, Return on Investment (ROI), Total Cost of Ownership (TCO)

1. INTRODUCTION

The fundamental goal of all Return on Investment (ROI) studies is to describe how existing customers derive value from the use of a particular software product, and to extrapolate a generalizable model from those experiences that can predict the degree to which a new user will derive a measurable ROI.

Several 2003 reports claimed the ROI from CRM implementations was dismal, with 8 out of 10 projects failing to deliver on ROI promises, and 50-70 percent typical project failure rates. Some of this year's reports are more optimistic, with about 52 percent of companies responding that their CRM initiatives generated an ROI between 51 percent and 500 percent, and 30 percent of respondents saying the return was greater than 501 percent.

Many newcomer companies must decide if they should begin with a CRM initiative. Above mentioned results should have a great influence on their decision. No matter how different the results are, the companies must know as exactly as possible if the CRM project is in the long time run lucrative for them.

Before CRM project companies should be able to demonstrate ROI on their CRM investments, and they need to implement serious yardstick work when seeking to evaluate CRM-software investments and to give a clear evaluation that quantifies the expected costs, tangible financial benefits, intangible strategic benefits and risks.

In the following, we are going to discuss thoroughly costs, benefits and ROI problems.

2. COST-BENEFIT ANALYSIS OF CRM IMPLEMENTATION

In order to establish a business case for pre-project planning, and post-project success measurement, we must assess implementation costs, benefits, and risk.

i. Implementation Costs

Basically, implementation costs are should include CRM software licensing and maintenance or support contracts; EDI, database, operating system and other software licensing and maintenance or support contracts; hardware purchases and maintenance or support contracts including servers, storage and network upgrades (as required); software integration and customization, including design, development, test and ongoing maintenance; implementation labor; ongoing administration and support labor; planning and requirements meetings; user training and learning time; and process change management.

ii. Tangible and Intangible Benefits

As table 1 described, generally speaking, benefits typically include increases in staff productivity, cost avoidance, increased revenue and margin, and reduced inventory through the elimination of errors. Benefits could be divided into tangible and intangible benefits.

* Sponsored by Special Scientific Research Fund for Doctoral Education Base of Higher School (20030614011), National Science Fund of Excellent Youth(79725002)

Table 1 CRM benefits and improvements

Key Benefit Area	Key Improvement
Reduce cost of sales	Enable self-service sales channels, changing the mix of direct sales, tele-sales, channel or agent sales, and self-service the most effective or least costly channel in order to reduce costs and improve satisfaction
Reduce sales administrative overhead	Reduce the time spent on sales administrative overhead tasks such as commission calculations, forecasting and reporting enabling increased selling time
Improve leads to sales closure rates	Increase the percentage of leads which are converted to sales
Increase customer retention	Reduce customer churn rate and eliminate replacement expenses
Improve customer satisfaction and loyalty	Improve customer lifetime value

1) Tangible Net Benefits

A clear and precise cost-benefit analysis which tallies all of the planned project costs quantifies each of the tangible benefits and calculates key financial performance metrics such as ROI, NPV, and payback period.

2) Intangible Benefits

A total of the expected intangible benefits should include key performance indicators (KPI) that will be used to measure success or shortfalls of CRM project. These are a handful of areas of improvement that should be considered as Table 1 described.

3) Risk Assessment

Risk Assessment is also essential for CRM project success, because the evaluating the risks of people, process, and technology can proactively mitigate their probability and manage their impact on project success.

3. TCO AND ROI ANALYSIS OF CRM PROJECT

A survey of trends in the evaluation and purchase of CRM systems, released by Sage CRM Solutions, indicates that return on investment, cost and functionality are the three top issues for companies in the small to mid-market. Integration also figures significantly. However, ROI ranked as most important issue when evaluating CRM solutions.

We can get such conclusion, that ROI is the ultimate indicator of the success of a CRM program.

i. The Control of TCO

The first benefit is Total Cost of Ownership (TCO). When implementing a CRM technology, the initial costs of licensing, hardware, software, training and consulting, along with the ongoing costs of support and maintenance, all play a role. In the past the failure to account for all of the implementation costs helped lead to CRM failures. Today, smart companies are working with vendors to keep

TCO to a minimum. The lower the TCO, the quicker a company can begin realizing ROI on its CRM investment.

ii. The Enhancement of ROI

ROI from CRM typically comes in two forms: the first is cost reductions from increasing efficiency and the second form of ROI is revenue enhancements.

Through segmenting the different management layers, the main factors which have great influence on the ROI are described in the following text.

1) Strategic layer

The way to realize lower TCO and higher ROI should include the following strategic imperative.

- Develop overall strategic plan that leverages the customer base as a valuable asset and grows that asset over time.
- Differentiate customers according to their value to the firm and their individual needs, and set a plan to focus resources on top customers.
- Provide strong executive leadership through ongoing visibility and reinforcement of customer-focused goals and rewards.
- Commit to organizational change across the enterprise in order to communicate strategic objectives, spark adoption and instill processes.

2) Technology layer

Technology selection and framework will determine the success or failure of the CRM project, so that technology imperatives should be as following:

- Create a cross-functional team to define technology needs up front, prior to investment.
- Establish a strong relationship with a credible vendor capable of meeting your business needs and goals while keeping the initiative on time and on budget.
- Develop an integrated, scalable technology framework able to grow with the company over time.
- Match technology to processes built around the customer in order to empower customer facing-personnel and drive interconnected business long term.

3) Analytical Layer

a) Cost reductions

Analytical CRM solutions decrease costs by enabling the organization to:

- Target campaigns more accurately to tightly defined customer segments instead of large, scattershot efforts.
- Reduce customer support requirements by offering products/services that better suit their needs.
- Eliminate products and services that fail to meet customer expectations.
- Reduce the cost of acquiring new customers, by getting more bites for the buck on marketing efforts and doing a better job of retaining customers.
- Tailor interactions differently to high-value and low-value customers by focusing more expensive communication channels and/or more attractive offers

where the payback is higher.

b) Revenue Enhancements

Analytical CRM solutions increase revenue by enabling the organization to:

- Cross-sell and up-sell more effectively. It is easier and less costly to retain existing customers than to attract new ones; therefore, organizations can realize greater profits by increasing revenue from existing customers rather than relying on all revenue growth to come from new customers.
- Predict which customers are most likely to buy, by recognizing that customer needs and wants are constantly changing based on multiple factors, such as life transitions, major milestones or acquisitions, external events and more.
- Identify high-value customers and cultivate long-term relationships, by focusing on the top 5 percent of the customer base that very well might account for 75 percent of revenues, and in the process, increasing the percentage of high-value customers.
- Increase brand awareness by providing well-targeted communications that will be well received, and by providing consistent customer treatment — a uniform company image — across all customer touch points.
- Increase customer satisfaction, loyalty and referrals by using the customer insights generated by analytical CRM to differentiate the organization from the competition with customer-pleasing actions, offers, products and services.

4) Operational layer

a) Cost Reductions

In the following field we could realize the cost reductions:

- Service and Contact Center: reduced talk time, reduced post-call handling time, fewer call transfers, reduced wait times
- Sales: shorter sales cycle time, reduced cost of sales, less down time
- Marketing: reduced number of campaigns to cold prospects, fewer duplicate records and reduced levels of inaccurate customer data

b) Revenue Enhancements

In the following field we could enhance the revenue:

- Service and Contact Center: improved cross-sell and up-sell rates, higher problem resolution, higher customer satisfaction and re-purchase
- Sales: higher lead-to-close ratio, increased margins, more targeted and productive visits
- Marketing: improved response rates, increased number of annual campaigns, higher number of quality leads, increased average order size.

4. CAUSES OF HIGH FAILURE RATE OF CRM

Here we summarized some important reasons caused high failure rate of CRM implementation.

i. Strategic Problem

1) Implementing before creating strategy

A surprising number of firms dive headlong into CRM without carefully working out an exact customer strategy that the technology helps the company execute. Before even considering a specific CRM technology, organizations should figure out where they are going. This would include such basic planning actions as segmenting customers from most to least profitable, deciding whether to simply invest more in profitable customers, manage costs better to improve overall margins, divest unprofitable customers, or a combination of various options.

2) No long-term strategy

Believing that CRM is a technology solution is still a tremendous obstacle for far too many firms. The fact remains that CRM is a business process change, often supported by technology. But there continues to be a tendency to look to technology as a sort of business panacea. Business leaders who do so are often disillusioned by CRM, because they don't align their business processes to meet specific goals.

3) Lack of guidance

Companies would never construct their offices without a blueprint. According to Gartner, however, more than 60 percent of companies that have implemented CRM did not have mutually agreed upon goals for their projects prior to the installation. Like a building without a bearing wall, a CRM initiative without goals will collapse.

4) No risk assessment or contingency plan

ii. Organizational And Management Problem

1) Lack of communication between IT and the business, resulting in a mismatch of requirements and expectations

2) No senior business sponsor AND separate project manager

3) Lack of employee support

It's natural to resist change. Top salespeople may ask, for example, why should we be forced to change our working habits, when those very habits helped us become so successful? On the other hand, poorer performers may fear the outcome of their managers having a window into their bad habits. Failure to convince these and other employees of the benefits of CRM often results in passive resistance and low employee-adoption rates.

4) Technology put before people

No or minimal involvement of key users during the scoping phase and lack of regular communication with them throughout the project implementation.

5) No accountability

Driven by fear of the unknown, resistance also spills into the managerial level in the form of avoidance, or lack of accountability. But CRM success depends on that accountability. The fact is, if accountability is not taken at the upper most managerial level, negligence there will only breed negligence among those who should be using the system, but aren't, making failure inevitable.

6) Organizational culture is not aligned with the customer

Often management gives lip service to the notion of customer as king, but doesn't act that way. Workers are told to "delight the customer" but are actually rewarded for pushing peak volumes of product out the door, without regard for customer satisfaction. This type of company tends to overemphasize product metrics — such as inventory, costs and profitability — instead of customer measures, such as customer satisfaction, complaint levels, response times, loyalty, attrition and lifetime value.

7) Restricted marketing scope

The current CRM practice restricts its marketing scope within a company's database. These companies concern more and more about the data in their databases, but pay less and less attentions to their market competitions and marketing potentials, especially the potential customers NOT yet in their databases. These companies lost their visions to their markets and their marketing basically becomes a sub-function of their IT department.

iii. Customer Process Problem

1) Installing CRM in an organization that isn't already customer focused

It is no wonder a large percentage of CRM failures stem from insufficient change management based on trying to force CRM down the throats of staff that don't yet have a customer focus or an organization that's processes are not conducive to the building of strong customer relationships. It is folly to embark on CRM without taking the time to survey the company then reconfigure all processes and systems so they actually fulfill customer needs.

2) Incorrect definition of customer relationship

The current CRM practice incorrectly defines customer purchases and contacts as customer relationship. But customer purchases and contacts are the tangible customer behaviors rather than the customer relationship that drives these behaviors. Until it understands the driving force behind customer loyalty behaviors, a company is unlikely to be able to manage its customer relationship properly.

3) Confusing customer interaction with hounding

In some cases, organizations have implemented CRM in such a way as to hound customers with repeat messages by phone, fax, email or direct mail. While marketing has used CRM to devise innovative methods of inundating the mailing list, this approach may actually drive business

away in the long run. The important thing is to identify the right customers i.e. those that actually want to form a relationship with you. Next, you have to work out how to contact them in the best way so that they continue to value the relationship.

4) Ignorance of customers' needs

The current CRM practice lures companies away from their markets and their customers. The current CRM practice requires companies to focus on things like the correlations between product purchases for cross/up selling and the predictions of what customers will buy in the future from the products/services in their databases. So the companies that adopted the CRM practice tend to ignore their customers' real needs, which are the real drivers behind the purchase correlations.

iv. Project Problem

1) Poor project scoping and undefined project objectives, roles and responsibilities

These will lead to the setting of unrealistic expectations.

2) No project success metrics

Without project success metrics CRM implementer can't efficiently monitor and evaluate the project under construction, so that enterprises are not able to guarantee the success of CRM projects.

3) Lack of regular checks to ensure the project is on track - to time and budget

The direct aftermath of lack of regular checks to ensure the project is on track - to time and budget is that implementer of CRM can't strictly control the end time of accomplishment of CRM project with acceptable budget.

v. Technology Problem

1) The technology gap

Three key limitations typically undermine systems that support analytical CRM strategies:

- Organizations don't have an enterprise-wide view that reconciles and exploits data from all relevant functional/transactional systems across business units.
- Organizations may have simple analytics that tell them what was and what is, but not an accurate view of what will be, why and what to do about it.

2) Solving CRM implementation problems with more CRM technology

Some companies try to implement a state-of-the-art CRM system at once. They spend, and often waste, millions in doing so. Their faulty reasoning considers that the latest technology will instantly solve a decade's worth of people problems that plague the business. More often than not, the expensive technology goes largely unutilized and the same problems continue to haunt the company. In many cases, then, it may be better to start small with a relatively

low-tech approach to CRM.

3) Integration Woes

Today there is no killer application that solves all integration problems. Most large-scale implementations require some customization. This may lead to problems that put vendors and consultants at odds with customers.

Many vendors and consultants maintain that most customers expect integration to happen like plugging a light fixture into a socket and flipping a switch, when in fact it is an evolutionary process.

4) Dirty Data

An often-overlooked, yet insidious hurdle is dirty data, or inaccurate and old information. Data is the lifeblood of a CRM system, and incorrect numbers, spelling mistakes, and outdated contact information can infect that system if it is left unchecked.

5. SUCCESS FACTORS IN CRM

The benchmarking consortium detailed and prioritized measurements for each benchmark identified in the kickoff meeting and guided the entire selection of successful practices. Our analysis led to the formulation of six success factors.

1) Evolution Path

As a first step, most companies implemented a system for operational CRM, e.g. call center or sales force processes. Analytical CRM rests upon this foundation and covers data mining and churn analysis functionality. Only little use was made of collaborative CRM which indicates an evolution path from operational, analytical to collaborative CRM and a stepwise implementation of CRM.

2) Timeframe

Most successful practice organizations have gone through a rapid system evaluation phase and completed the system introduction phase within approximately 7 months. But a full CRM was considered to take a minimum of 2 years. Successful companies split these long-term CRM projects into manageable subprojects lasting a maximum of 6 months.

3) Organizational Redesign

Prior to the introduction of a CRM system all successful practices established CRM concepts for the definition of processes and organizational structures. Organizational redesign also needs to consider structural issues, i.e. to establish a centralized responsibility and authority for defining cross-functional standards.

4) System Architecture

Virtually all the CRM systems in the benchmarking were standard packages, while no system offered a comprehensive operational, analytical and collaborative CRM functionality. Advanced CRM companies integrated specialized systems for operational, analytical, and collaborative CRM into a best-of-breed architecture.

Following the evolution path described above, mature CRM concepts also required integration architecture for seamlessly exchanging information.

5) Change Management

Convincing employees of the benefits of CRM methods and systems is regarded as an important success factor.

6) Top Management Support

For establishing customer orientation on a corporate level, implementing inter-organizational process and system standards, and for supporting the adoption of the CRM systems within the organization top management support was a key requirement. Top management sponsors also ensure that short-term setbacks in the CRM project can be overcome.

6. CONCLUSION

In order to do an ROI study correctly, we need enough customers with enough experience. What constitutes enough customers is the first problem. The validity of the sample is the next problem. The third problem is that even if you get a lot of ROI data from a lot of customers, it's rarely in a form that can be used to build a generalizable model. Every company had a different way to measure ROI.

Because of above-mentioned problems we have not tried to construct a generalizable panacea model, but only proposed criteria and critical factors to take account of during evaluating the RIO of CRM implementation of enterprise. There is very close linkage between the enhancement of ROI and the success rate of CRM implementation. If we pay enough attentions to success factors in CRM and avoiding causes of high failure rate of CRM projects, the ideal level of ROI must be reachable.

ACKNOWLEDGEMENT

ICEB2004 is supported by the National Natural Science Foundation of China.

REFERENCES

- [1] Francis Buttle. *Customer Relationship Management- Concepts and Tools*. New York etc. ELSEVIER Butterworth Heinemann, 2004
- [2] Federico Rajola. *Customer Relationship Management -Organizational and Technological Perspectives*. Springer -Verlag, Berlin etc, 2003
- [3] David Myron. *Barriers to CRM Success And How to Overcome Them*. CRM Magazine, August 2003
- [4] SAS Institute Inc. *Maximizing ROI from CRM Initiatives*. A SAS White Paper, 2002
- [5] Drew Robb. *The Four Perils of CRM and How to Avoid Them*, October 25, 2002
- [6] Dick Lee, David Mangen, Bob Thompson. *The Blueprint for CRM Success*. HYM Press, 2003
- [7] Joshua Greenbaum. *The Paradox of ROI*. http://www.intelligententerprise.com/021115/518enterprise1_1.jhtml?_requestid=186794

Comparing Sales Strategies Using the Markov Chain Relationship Model

Bocheng Chen¹, WH Ip², Peng Sheng³, HL Yu²

1.School of Economics & Management, Tsinghua University, Beijing 100084, China,

2.Department of Industrial and Systems Engineering, The Hong Kong Polytechnic Univ, China,

3.Beijing Time Sonic Consulting Co., Ltd, Beijing 100085, China,

1.chenbch@em.tsinghua.edu.cn; 2.mfwhip@inet.polyu.edu.hk; 3.sp@5sbts.com

ABSTRACT

In this paper, the author applied the concept of the Markov chain and divided sales procedures into several indexes and states; use the state index for connecting success in sales and customer relations into Pfeifer's method, establish a mathematical model, and demonstrate its result. In order to increase profits and decrease the cost of sales for the company, we further classify customers and propose different sale strategies. Case study and analysis are provided to elaborate the approach and its contribution to sales and CRM (customer relationship management) strategy.

Keywords: CRM, Relationship model, Selling process, Markov Chain

1. INTRODUCTION

Customer relationship modeling is always a popular topic for research. Sheng Peng conducted a thorough analysis, from the aspect of psychology, on the "customer person" in the process of selling.

He^[1] considered that the selling is sales to people- to key customer people (KP, who are responsible for or have a direct controlling relation with the purchasing), no matter to a plant or a family. In the case, a selling is involved several factors, besides the enterprise's "selling points", customer's "purchasing point", more important, a good salesperson should also supply a KP with some "selling points" in KP's group, those selling points should be not only accepted by the KP from his heart, but also can be accepted by KP's group members, even more, some of the key ideas may let the KP believe the purchasing can bring him more respect in his group, and the KP is willing to declare the purchasing suggestion in the public. In such cases, the selling will be success in very large probability. He also thought that a sales process should be able to measured, it should never be managed in a "black box", managers can't handle the sales process and know the result only when it is out from the other side of the box. Such things happen because there is no state index set in the sales process, Sheng set up three indexes to help solving the problem.

Also to solve the problem, Reichheld, F.F.^[6] consider that only one index is needed: "You simply need to know what your customers tell their friends about you". Morgan^[7] and Kristensen^[8] do not agree with his idea.

With regard to the customer relations mathematic model and the aspect of analysis, though many papers involving in the topic, most of them are methods based on the contacting and sales result information analyzing, only a few of them are modeling inside the process. Blattberg and Deighton^[2] proposed an LTV(Life Cycle Value) model to compute the cost of acquiring and

maintaining customers. In order to help managers to optimize their sales, Bronnenberg^[3] made an attempt to model and analyze consumer relationships using the Markov chain (MC); Based on supposed several states, Pfeifer and Carraway^[4] constructed several CRM models with MC, and computed and analyzed different sales strategies within given parameters with profit as the objective, but they do not given more detailed computing and analysis to their model, also omitted that the supposed states can help salespeople improve their work; Jain and Singh^[5] conducted a thorough survey of the area, describing the current and future trends.

In this paper we will introduce Sheng's indexes and apply the indexes into Pfeifer's^[4] Markov model, then adopt different sales strategies, and a little detailed computing, to test the model, and analyze the benefit to the enterprise. In the analysis the customer classifying method will be employed, the results computed, and various scenarios compared.

2. THE SALES STATE INDEX, SELLING PROCESS, KNOWLEDGE, AND MARKOV MODEL

2.1 The Sales State Index, Selling Process Classification and Knowledge Management

In order to measure a sales process, Sheng^[1] set up three index sets, based on the relationship between salesperson and KP, from connection relation, attitude to product, and confidence to product sides. Each index and states are given as following:

Connection index (To customer): Has nothing to say (low), and only says the official words (middle), the conversation is valid (good), and no secrets are kept from each other (high);

Attitude index (To product): basically not approve the product(low), basically approve the product(middle), and approve the product very much (high)

Confidence index (To product): approval given on a

case-by-case basis (low), approval given with appropriate examination (middle), approval given in big or important situations (high).

Such indexes are index with “operating concept”, they can not only help salespeople make sure their sales state, but also supply them with an improved direction. For example, if your state is three “high” in the three sides, the selling will be success in very large probability; If your state is three “middle”, it is hard to say your sales result, what you should do is to improve your relation with KP, in the three given sides, in order to get better state and have larger sales success probability; If your state is three “low”, you can get success, but in a small probability.

Not as a traditional CRM system, the system built with such concept aims the data on KP’s information, customer purchasing point (PP), sales point in his group (SP), customer views on criteria (VOC) and so on. Such indexes classify a whole sales process into states, and its operating concept about states normalize salespeople’s sales state description and unify salespeople’s understanding to each given sale state; make sales managers monitor his people selling easily; locate the state and direct the further improvement direction, and so on. Under such concept, we notice that a customer is not just a customer, but a separated decision process. Its system is a special knowledge management system, what it stores is “human” intelligence and analysis result, not just sales results, because its unified state description makes such management possible. With such system, enterprise managers and researchers can classify customers not only from sales result data, but also from the KP’s character, PP, SP, and VOC, they can get more deeper and wider customer information, and what they get is the reason, not a statistical estimation, why customer buy or not buy their product. So such system can help the enterprise manage, share, accumulate, its salespeople’s sales knowledge and key customer information. We think that a system built with such concept should be a real customer knowledge management system (more detail information can be seen from Sheng^[1]).

2.2 Markov Chain Modeling

Based on the concept given above, we can discuss several sales strategies, when we separated a sales process in to several states, each with its sales success probability. Which scenario will be better, selling in large probability or selling in any chance? Following is the discussion by means of Pfeifer’s MC model^[4]. In order to simplify our discussion, we enlarge one of those index sets and use it to model the state of a sales

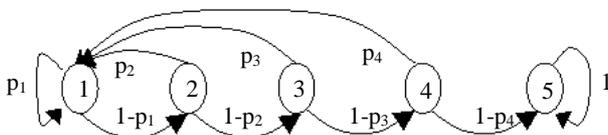


Fig 1 A sale procedure with given sale states

process. The confidence index set divides a sales process into the following five states: 5(former customer), 4(no approval in one-on-one situations), 3(approval in one-on-one situations), 2(approval in small catcall situations), 1 (approval in big or important situations). Obviously, state 1 holds the greatest probability of success. These relation states within a sales process are shown in Fig.1.

Pfeifer’s model^[4] is a model of the relationship between the sales and its customers (we set states following the thought). When the trading began, the relationship between the salesperson and the customer can be in any state in Fig.1, and the sale can be successful in any state, but the probability is different. Suppose the income of a successful sale is N=40, and the cost of the sale is 4. The probability of one sale succeed is the corresponding p_i. If the sale succeeds, the enterprise will obtain the benefit of N-M; if not, the enterprise will lose M, and the probability of the relationship going back one state is 1-p_i. Obviously, p_i < p_{i-1}, p_j=1, i=1,2, ...,j, with j as the state number, state 5 indicating no connection. Suppose also that the cost of the sale is entered just before the deal is made.

The matrix P is a one-step transition matrix, R is a reward vector, V is the expected net present value^[4].

$$P = \begin{bmatrix} p_1 & 1-p_1 & 0 & 0 & 0 \\ p_2 & 0 & 1-p_2 & 0 & 0 \\ p_3 & 0 & 0 & 1-p_3 & 0 \\ p_4 & 0 & 0 & 0 & 1-p_4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad R = \begin{bmatrix} N-M \\ -M \\ M \\ -M \\ 0 \end{bmatrix} \quad V = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ 0 \end{bmatrix}$$

Here M1=M2=M3=M4=M. In fact, the element of probability in the transition matrix (including in the result matrix P², P³, P⁴, etc.) represents the corresponding relationship between enterprise and customer after one or more sales. If required, we can set some threshold value or make hypothesis test to those elements which alert their relationship in case a problem occurs or the relationship is broken.

To be more precise, suppose that there is a discount rate to the present value after each purchase. Here, the discount rate is d=0.2. Thus, the expected net present value vector in j time of the sale, is(1):

$$V^j = \sum_{i=0}^j [(1+d)^i P^i R]^T = I * R + [(1+d)P]R + [(1+d)^2 P^2 R] + ... + [(1+d)^j P^j R]^T$$

In order to simplify the results, we often use an infinite horizon to describe the purchase. From (1), we can get the expected net present value as follows:

$$V = \lim_{j \rightarrow \infty} V^j = [I - (1+d)^{-1} P]^{-1} R \tag{2}$$

Supposed i=4, p₁=0.3, p₂=0.2, p₃=0.15, p₄=0.05, then:

$$V^4 = [50.115 \quad 4.22 \quad 0.592 \quad -1.98 \quad 0]^T$$

$$V = [52.32 \quad 5.554 \quad 1.251 \quad -1.82 \quad 0]^T \tag{2'}$$

The result shows that, in this case, 25% customer (in state 1) makes the 90% profit of the enterprise. The negative expected value -1.98, in state 4, means sales enterprise shouldn’t sell in such customer relationship,

because the success probability is too small. How to solve the problem? Classifying customer is one way to improve the condition.

3. CLASSIFYING ANALYSIS

The purpose of the analysis is to determine the profit of the sale enterprise, and the object of the research is to see the effect produced by different sale strategies adopted to different customers, after having divided customers into different types. In many cases, the sale enterprise does not hope to gain much from the first or second trade, but hopes to make money over a given period of trading. Therefore, for ease of analysis, we assume an infinite case for the sale. In this case, the negative value in the V vector of equation 2 means that if the first connection with the customer is at the corresponding state, no profits can be obtained from the sale even if the customer makes an infinite number of purchases. This is because the success rate in the state is too small or the sales cost is too high, relative to its income. We also transform equation 2 into 3 in the following analysis, to avoid the computation of d and the inverse operation in equation (2). We then have:

$$[I - (1-d)P] * V = R \tag{3}$$

Though there are 5 states in Fig.1, from the perspective of sales, there are only two kinds of customers: relationship and no relationship customers (or former customer ^[4]).

3.1 Sales Strategies for Two Kinds of Customers

3.1.1 The General Result for the Two Kinds of Customers Scenario

From equation (3), the sale enterprise's expected net present value equation is:

$$\begin{bmatrix} (1-\frac{p_1}{1+d})v_1 - (\frac{1-p_1}{1+d})v_2 \\ (\frac{1-p_2}{1+d})v_1 + v_2 - (\frac{1-p_2}{1+d})v_3 \\ (\frac{1-p_3}{1+d})v_1 + v_3 - (\frac{1-p_3}{1+d})v_4 \\ (\frac{1-p_4}{1+d})v_1 + v_4 \\ 0 \end{bmatrix} = \begin{bmatrix} N-M \\ -M \\ -M \\ -M \\ 0 \end{bmatrix} \tag{4}$$

then

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} v_1 \\ (\frac{p_2}{1+d})v_1 - M + (\frac{1-p_2}{1+d})v_3 \\ (\frac{p_3}{1+d})v_1 - M + (\frac{1-p_3}{1+d})v_4 \\ (\frac{p_4}{1+d})v_1 - M \end{bmatrix} \tag{5}$$

the result :

$$v_1 = \frac{N - (1-q_1q_2q_3)M}{1-p_1q_1 - p_2q_2 - p_3q_3 - 1} \tag{6}$$

therefore :

$$v_4 = \frac{Np_4 - M(1+q_1q_2q_3)p_4}{1-p_1q_1 - p_2q_2 - p_3q_3 - 1} - M$$

among them :

$$p_i = \frac{p_i}{1+d}, q_i = \frac{1-p_i}{1+d}, p_i q_i = \frac{1}{1+d} (i=1,2,3,4).$$

z From equation (4) and (2'), we notice that the response of customers in different states to same input M is very different, $p_1=0.3$, and $p_4=0.05$. From Markov chain's concept, we can see customers in state4 as one time purchasing customers (OTPC), and ones in state1 as familiar ones. In order to compare easily, supposing we can increase sales cost to promote OTPC's success rate, to think it as linear relation simply, $M_4=(p_1/p_4)M$, the result shows, in one side, the cost to acquire a new customer is as much as (p_1/p_4) times to retain a familiar one.

z From equation (5), we can get: $v_1 > v_2 > v_3 > v_4$.
 For: $v_4 = p_4 * v_1 - M$;
 $v_3 = (p_3 * v_1 - M) + q_3 * v_4 > (1+q_3) * v_4$;
 $v_2 = (p_2 * v_1 - M) + q_2 * v_3 > v_3$
 $v_2 < (p_2 * v_1 - M) + q_2 * v_1 = (p_2 + q_2) * v_1 - M < v_1$
 The v_2 's first item is $p_2 * v_1$, so $v_1 >> v_2$ and v_3, v_4 , when p_2 is small.

That means, under the given suppose, the better the relation is, the more the selling gets. The best relation customer will supply most of the sales profit.

z If the aim is to obtain greater benefits, $\sum v_i = \max$, at least, each $v_i \geq 0$ ($i=1,2,3,4$). From (5), each v_i is related to its p_i , and in this case, p_4 is the least, so if $v_4 \geq 0$, then other $v_i \geq 0$ ($i=1,2,3$). From (6), we can get the threshold p_4 and M as following:

$$p_4 \geq \frac{M(1-p_1q_1 - p_2q_2 - p_3q_3)}{NM(1+q_1q_2q_3)}$$

$$M \leq \frac{Np_4}{(1+q_1q_2q_3)p_4 + (1-p_1q_1 - p_2q_2 - p_3q_3)}$$

The Fig2 shows the relation between them, to the given data, the threshold p_4 in (2') should be $p_4 = p_4 * d = 0.09$, so in (2'), $v_4 < 0$. By the same way, we can get each threshold p_i and M ($i=1,2,3$).

3.1.2 The Modified Strategy 1 for the Two Kinds of Customers Scenario

Under the given condition, Sales enterprise should modify its sales strategy to avoid losing money. The

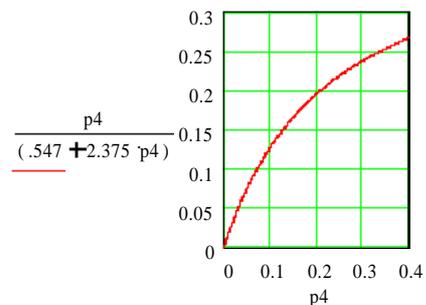


Fig 2 The relation between threshold p_4 and M modified strategy 1 is to take no sales action to custom-

ers in state 4, and treat them as no relation customers.

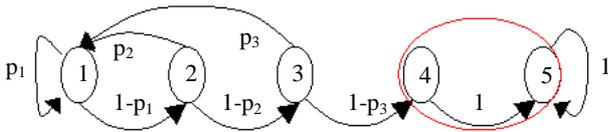


Fig. 3 The state figure of Scenario (7)

When $p_4 < M/v_1$, in formula (5) $v_4 < 0$; i.e. sales to customers in state 4 (no approval in one-on-one cases) can never yield a profit. So stopping sales to them and treating them as customers in state 5 is a new sales strategy (see Fig.3, $p_4=0$, the corresponding $M=0$, $V_4=0$). Putting those results to formula (5), we have:

$$V = [v_1 \quad p_2 * v_1 - M + q_2 * v_3 \quad p_3 * v_1 - M \quad 0 \quad 0]^T \quad (7)$$

 Compared with formula (5), there is no negative v_4 in v_3 , which is in (7), so v_3 in scenario (7) is better than that in (5); for the same reason, the v_2 in scenario (7) is also better than that in (5). This means that, whether single v_i or total profit($\sum v_i$) is involved, scenario (7) is better than that scenario (5). If $v_4 > 0$, the conclusion is different. Considered the (2') data, we have the result: $V = [53.149 \quad 6.621 \quad 2.644 \quad 0]^T$, it's better than (2').

The result means that enterprise should not sell in the very small success probability case.

3.1.3 Improved Strategy 2 for Two Kinds of Customers

Furthermore, we can take no sales action for customers in state 3,4, and to treat them as no relation customers, as our strategy 2.

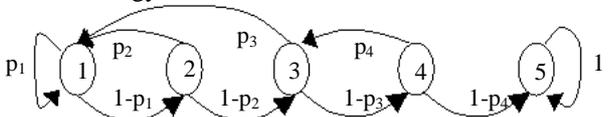


Fig. 4 The state in improved Scenario (9)

Similar to the previous analysis (Fig.3, also suppose $p_3=0$, corresponding $M=0$), taking the condition in formula (5), we get an improved scenario. In this case, the solution is:

$$V = [v_1 \quad p_2 * v_1 - M \quad 0 \quad 0]^T \quad (8)$$

Compared with formula (7), the scenario is better than that of (7) only when $v_3 < 0$. Otherwise, it will be worse than scenario (7), because, in this case, v_3 can obtain profit. To the data given in (2'), we have result: $V = [51.574 \quad 4.596 \quad 0 \quad 0]^T$, the result is not as good as (7), for the p_3 in scenario (7) is great its threshold p_3 .

From the above analysis, we can see that the sales strategy is available, because the sales enterprise can gain greater benefits, in the case of its environment not changed.

3.2 The Sales Strategy for Multiple Types of Customers

The case in Fig.3 is not realistic enough, because the enterprise should put some money into improving its

relationships with its customer and bring it to a higher state: when the enterprise has made too few successful sales. Therefore, in order to reduce sales costs, we can divide both customers and sales costs into three types: sales cost M , retain cost $H(<M)$, and 0.

3.2.1 Sales Strategy 1 for Three Kinds of Customers

A lower cost H is used for the case of improving customer relationships in order to reduce costs. Customers are classified into three types, such as 1,2,3,4; and 5; and the sales cost for state 4 is H (Fig.4). The difference with that of earlier is that current p_4 is much great than before, because the current p_4 is not the probability of making a successful sale, but the state of increasing one. v_i ($i=1,2,3,4$) is still the present net value of each state. From equation (3), we can obtain the sale enterprise's expected net present value equation for the case of infinite purchases, as follows:

$$\begin{cases} v_1 = p_1 N_1 (M)^{-1} - (1-p_1)M \\ v_2 = p_2 N_2 (M)^{-1} - (1-p_2)M \\ v_3 = p_3 N_3 (M)^{-1} - (1-p_3)M \\ v_4 = p_4 (v_3 - H)^{-1} - (1-p_4)*0 \end{cases}$$

The transition matrix P and reward vector is:

$$P = \begin{bmatrix} p_1 & 0 & 0 & 0 & 1-p_1 & 0 & 0 & 0 \\ 0 & p_2 & 0 & 0 & 0 & 1-p_2 & 0 & 0 \\ 0 & 0 & p_3 & 0 & 0 & 0 & 1-p_3 & 0 \\ 0 & 0 & 0 & p_4 & 0 & 0 & 0 & 1-p_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad R = \begin{bmatrix} N-M \\ N-M \\ N-M \\ v_3-H \\ -M \\ -M \\ -M \\ 0 \end{bmatrix}$$

From equation (3), we obtain the expected net present value as follows. The symbols are the same as those in

$$V = \begin{bmatrix} \frac{(N-M)M}{1-p_1} & \frac{(N-M)q_2M}{1-p_2} & \frac{(N-M)q_3M}{1-p_3} & \frac{v_3-H}{1-p_4} & 0 \end{bmatrix}^T \quad (9)$$

(6). Because formula (7) can be shown as following:

$$V = [v_1 \quad (p_2+q_2*p_3)v_1 - (1-p_2)M \quad p_3*v_1 - M \quad 0]^T$$

and to prove which is larger((7) or (9)) is a little hard,

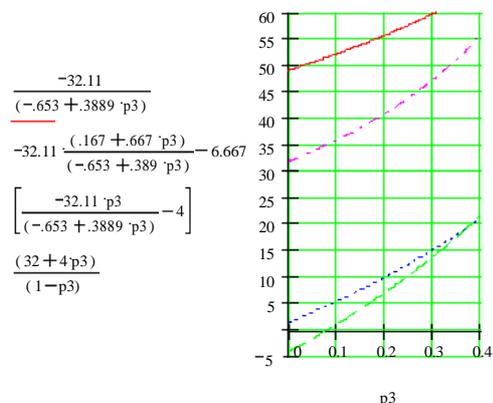


Fig.5 Comparing scenario (7) with scenario (9)

so we can deal with it by Fig.5 (with the data in (2')), the first three formulas are v_1, v_2, v_3 in (7), and the fourth one(the line in the Fig middle) is the v_3 in

scenario (9), the other two is: $v_1 = 44.89, v_2 = 40$, when p_3 varying from 0 to 0.4, so $\sum v_i$ in scenario (9) is much better than that $\sum v_i$ in scenario (7) (note: we haven't consider the v_4 in scenario (9) in the analysis). $V = [44.89 \ 40.00 \ 37.90 \ 49.47 \ 0]^T$ is the (9) result with the data in (2'), except $p_4 = 0.3/1.2$.

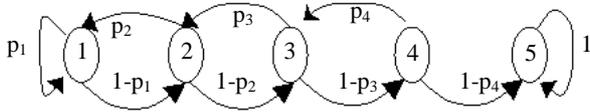


Fig. 6 The state in an improved Scenario (10)

The result means that enterprise should improve the relationship with customer first, and reduce the sales chance under too low success probability case.

3.2.2 Characteristics of the Transition Matrix

In a transition matrix with structure as such P, all customers will finally become no connection customers. Because when $p_{ij} \geq 0$ in the matrix P, there must be a steady limit probability vector. Suppose this vector is W and that $W = (W_1, W_2, W_3, W_4, W_5)^T$. We then have a steady limit probability equation $W = P^T * W$. For the character of matrix P, we obtain the first two lines as:

$$\begin{cases} p_1 W_1 + p_2 W_2 = W_1 \\ (1 - p_1) W_1 = W_2 \end{cases}$$

we have: $\begin{cases} p_2 W_2 = (1 - p_1) W_1 \\ W_2 = (1 - p_1) W_1 \end{cases}$ Because $p_2 \geq 0$, then,

$$W_1 = W_2 = W_3 = W_4 = 0; \sum W_i = 1, \text{ so } W_5 = 1.$$

The result means that a transition matrix with such a structure will result in decreased relationship as the trading time increases; finally, all customers will become no connection customers (steady state). The Pfeifer's calculation of P, P², P⁴ also shows similar characters^[4].

3.2.3 Sales Strategy 2 to Three Kinds of Customers

We can improve our sale strategy further, and propose a more conservative sales strategy: the enterprise should only conduct its sales in state 1,2.

The sales enterprise may require a regulation that the product must be sold at the high relation state, i.e. where there is a high probability that the sale will be a success—in "approval in small catcall situations" and "approval in big or important situations." Therefore, by inputting H sales cost, each success in state 3,4 can increase the step of the state forward by one state; each failure can set the relationship back by one state; and the sales action can only occur in states 1 and 2 with a cost of M. This strategy is exactly like the one that classifies customers into the following three types: 1,2; 3,4; 5, only now Fig.5 is changed to Fig.6. This is a typical random walk model. Hence, we can establish the equation by using formula (3):

$$\begin{cases} v_1 = N(M - H)^{-1}(1 - p_1)M \\ v_2 = p_2(N - M)^{-1}(1 - p_2)M \\ v_3 = p_3(2 - H)^{-1}(1 - p_3)H \\ v_4 = p_4(3 - H)^{-1}(1 - p_4)H \end{cases}$$

The transition matrix P and reward vector are:

$$P = \begin{bmatrix} p_1 & 0 & 0 & 0 & 1-p_1 & 0 & 0 & 0 \\ 0 & p_2 & 0 & 0 & 0 & 1-p_2 & 0 & 0 \\ 0 & 0 & p_3 & 0 & 0 & 0 & 1-p_3 & 0 \\ 0 & 0 & 0 & p_4 & 0 & 0 & 0 & 1-p_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R = \begin{bmatrix} N - M_1 \\ N - M_2 \\ v_2 - H \\ v_3 - H \\ -M_1 \\ -M_2 \\ -H \\ 0 \end{bmatrix}$$

The expected net present value is equation (3) after the purchase (Suppose $M_1 = M_2 = M$).

$$V = \begin{bmatrix} (N-M)M \\ (N-M)q_2M \\ (v_2-H)q_3H \\ v_3H \\ 0 \end{bmatrix}^T \quad (10)$$

Comparing with scenario (9), there is no change in v_1, v_2 of scenario (10). The value of p_3, p_4 changes (the change will result in v_3 's numerator becoming larger and its denominator becoming smaller), but the change in v_3 results in a change in v_4 . If we can prove v_3 in (10) is better than v_3 in (9), then we can say scenario (10) is better than scenario (9). We have:

$$\frac{v_2 - (1+q_3)H}{1-p_3} - \frac{N - (1+q_3)M}{1-p_3} > \frac{v_2 - (1+q_3)M}{1-p_3} - \frac{N - (1+q_3)M}{1-p_3} = \frac{v_2 - N}{1-p_3}$$

Notice that, we use M instead of H for the comparing.

$$\frac{v_2 - N}{1-p_3} = \frac{N - (1+q_2)M - N}{1-p_3} = \frac{p_2 * N - (1+q_2)M}{(1-p_2)(1-p_3)}$$

When $\frac{p_2}{1+q_2} > \frac{M}{N}$ the v_3 in (10) is better than v_3 in (9).

Consider H, we can say that the scenario (10) is better than scenario (9). Also we can get result:

$$V = [44.89 \ 40.00 \ 51.64 \ 67.79 \ 0]^T$$

(Here: $p_4 = p_3 = 0.3/1.2$), it looks better than that in scenario (9).

4. CONCLUSION

In the paper, we analyzed the customer relationship, set up the sales state index for the sales process, and modeled the relationship by introducing the concept of the MC based on the given state index; and presented the results of our analysis of the model. Pfeifer^[4] pointed out that such an approach can also be applied to analyzing the problem of "Recency," "Frequency", and "Monetary."

The proposed approach here is to model the relationship with the MC by using various indexes, analyze the relationship between cost and the probability of making a successful sale in terms of benefits obtained, test different sales strategies on the model and present the corresponding solution.

The analysis results show us that the most conservative sales strategy is much better than the normal strategy.

That means, in the case of small sales success probability, the sales enterprise shouldn't sell in any state at any chance, but to improve the relationship with customer first and prepare to sell in the high success probability state.

ACKNOWLEDGEMENT

This study was partly supported by the Key Laboratory of Information Management and Information Economics, Ministry of Education P.R.C (F04-23).

REFERENCES

- [1] Sheng Peng. What is selling? [M]. The Social Science Publisher of China, 2003.4.
- [2] Blattberg, R., & Deighton, J. Manage Marketing by the Customer Equity Test [J]. Harvard Business Review, July/August, 1996:136–144.
- [3] Bronnenberg, B.J. Advertising Frequency Decisions in a Discrete Markov Process Under a Budget Constraint [J]. Journal of Marketing Research, 35 (3), 1998:399–406.
- [4] P Pfeifer, R Carraway. Modeling customer relationships as Markov chains[J]. Journal of Interactive Marketing, 14(2), 2000:43-55.
- [5] D Jain, S Singh .Customer lifetime value research in marketing: a review and future directions[J]. Journal of interactive marketing,16(2),2002:34- 46.
- [6] Reichheld, F.F. .The One Number You Need to Grow[J], Harvard Business Review, 81(12), Dec 2003.
- [7] Morgan, Neil A., Rego, Lopo L..The One Number You Need to Grow[J]. Harvard Business Review, 82(4), Apr2004
- [8] Kristensen, Kai1,Westlund, Anders2.The One Number You Need to Grow[J].Harvard Business Review, 82(4),Apr2004:136-137

Consumers' Trade-Offs of Online Auctions Alternatives -- An Application of Conjoint Analysis

Hector Mingta TSAI¹, Yichen LI²

¹ Graduate Institute of Management, Fo-Guang University, Ilan, Taiwan, China

² Graduate Institute of Informatics, Fo-Guang University, Ilan, Taiwan, China
{mmtsai; g2332001}@mail.fgu.edu.tw

ABSTRACT

As auctions buyers and sellers become proliferated on the Internet, there are considerable interests in investigating consumer behavior in the new marketing context. Much of the research on online consumer behavior has focused on the economic dynamics of bidding and offerings, technical mechanism and privacy, while consumer trade-offs between auctions alternatives are still lacking in literature. The research objective of this paper is to examine how consumers evaluate each seller's attributes and make bidding decisions. A conjoint analysis was used to investigate subject's bidding intention and trust for a variety of used textbook sellers that differed in terms of bidding price, seller rating (feedback score), payment type, and shipping charge. This article provides an experimental test of consumer's trade-offs in online auctions context. These findings are essential because they help identify the importance of those attributes of online auctions under the consideration of bidding intentions and seller's trustworthiness. Interestingly, this paper suggests auctions sellers may succeed by simply compete on price. However, this conclusion could only be generalized in used textbook category; further research could include additional product categories as exemplars and different auctions attributes as factors. Although price and seller rating seem to be relatively more important factors, further research could examine factors in addition to payment type and shipping charge of the auctions.

Keywords: trade-off, online auctions, conjoint analysis, internet marketing

1. INTRODUCTION

In the past few years, online auctions have been the most profitable business model on the web. According to statistics reported by A.C. Nielson, online auction bidders grew three times over the past two years in Taiwan. Take the famous Internet auction site, eBay and Yahoo! Auction, for example, people visit the online auction marketplace to buy and sell items in thousands of categories including clothes, antiques, books, computers, DVDs, real estates, electronics & computers, etc. With huge arrays of choices, online auctions have not only changed the parameters of traditional auctions, but also provided the sellers with more degrees of freedom. The Internet provides consumers with multitude choices and convenience, at the same time, risks and uncertainties on the web are inevitably major concerns for buyers and sellers due to online anonymity and fraud threats.

Over the years auctions have become known as not only economic events, but also as socially and culturally significant affairs. As auctions buyers and sellers become proliferated and active on the Internet, there are considerable interests in investigating consumer behavior in the new marketing context. However, most of the research on online auction behavior focused on the dynamics of bidding and offerings, auction platform mechanism and privacy, while consumer's decision making for online auctions alternatives are still lacking in literature.

The research objective of this paper is to examine how

consumers evaluate each seller's attributes and make bidding decisions. In this paper, we examine subjects' preferences for a variety of auction alternatives that differ in terms of four major attributes, which are hypothesized to be important to consumer's bidding judgments and perceptions of seller trustworthiness, and compare how demographics and web experiences influence consumer's preferences.

In the following sections, we review the benefits of online auctions, and consumer's evaluation criteria that are used in this research. Then we develop hypotheses to explicate the factors that affect consumer's trade-off between attributes. We use conjoint experimental design to test the hypotheses, and the results will be reported. In the final section our conclusions of the study will be provided.

2. ONLINE AUCTIONS

2.1 Benefits of Online Auctions

As economic scholar Steve Kaplan (1999) of Chicago University pointed out that online auctions have greater economic advantages as compared to traditional auctions. Hanson (2000) contended that online auction sites improve the power and efficiency of auctions, because: (1) the Internet makes it easier to gather buyers and sellers together in the same place at the same time; (2) the Internet enables sellers to provide in-depth information, so buyers can evaluate the item being sold; (3) the Internet expands the number of bidders, which raises the price paid and the profitability of the auction.

Online auctions are very appealing from the viewpoints of both businesses and consumers. The auctions are a typical example of the ideal online business: they act as intermediaries between buyers and sellers; except for office space, they do not require extensive investments in physical property as well as inventory. For consumers, the online auctions websites offer access to tens of thousands of various goods around the world. On the other hand, the competitive nature of the online auctions' bidding processes enables sellers to achieve a better price than they might have using traditional media such as newspaper classifieds.

2.2 Attributes of Online Auctions

Although online auctions have many advantages for both consumer and seller as mentioned above, there are disadvantages that are obstacles for consumer to participate in an auction deal. First, there is the potential for fraud from dishonest or irresponsible sellers, which could cause buyer losses. Second, there are also concerns of financial privacy issues and personal information misuses. Third, consumer may have to accept undesirable restrictions because of failing to read the fine print. Fourth, products may not always cheaper than a local store, especially when the cost of shipping is included.

To overcome the potential problems in online auctions, consumers have to evaluate the information given by sellers on the web, and make the decision. Our research goal is to examine how consumer tradeoff among different sellers under the consideration of bidding intention and seller trustworthiness for a given product. We select four attributes that are fundamental to consumer's decision making in evaluation of a seller under the context of online auctions. These attributes are price, seller rating, payment type, and shipping charges.

2.2.1 Price

The pricing mechanism is the most discussed issue in auctions. Increasing information on the web may lead to less willingness to pay. Common perception is that the Internet will raise consumer price sensitivity, especially when the websites enables instantaneous side-by-side price comparisons of available alternatives (Hanson, 2000). However, Preliminary research concerning the price of online items finds increased price dispersion (Brynjolfsson & Smith, 1999) with online prices often surpassing the price of similar products purchased through more traditional venues (Lee, 1997; Bailey, 1998). It remains to be seen if price competition significantly increases as electronic markets mature (Smith, Bailey & Brynjolfsson, 1999).

2.2.2 Seller rating

The most common type of consumer protection feature

offered by online auctions is a user rating service, in which buyers and sellers can rate each other based on how promptly product was received, how quickly a buyer provided payment, etc. These ratings are maintained on auctions' websites and are easily accessible by users. Auction participants can use feedback systems to publicly rate their satisfaction towards their trading partners. Specifically, the feedback system is a measure of a user's reputation in an auction community. The auctioneers encourage all users to check their trading partners' rating before transactions and leave feedback about their trading partners after their transactions. In essence, the system tries to use one's reputation as a deterrent for cheating behaviors. That is, if one develops a bad reputation, other auction participants may not transact with the person anymore (Ba, Whinston, and Zhang, 2003).

2.2.3 Payment type

There is a wide range of ways to pay on auctions websites, such as eBay or Yahoo, and most sellers offer more than one option. Some payment options offer more protection than others, so consumers are prompted to select a method of payment that makes them feel comfortable, especially when most of the auctions require buyers to pay in advance. There are three payment types that usually used in Taiwan, cash, bank transfer and credit card. In this study, we use these three options in the experimental design.

2.2.4 Shipping charge

For most of transactions on the Internet, payment and delivery rarely occur simultaneously. Notable exceptions exist where payment and delivery can occur nearly simultaneously for online purchases (e.g., when purchasing MP3 files). More commonly, the buyer is required to pay in advance for a product or service to be delivered sometime into the future, often with little or no ability to examine the product or service in advance. Specific shipping costs are noted in item descriptions approximately 50% of the time and almost every seller require buyers to pay for shipping and insurance fees (James H. Gilkeson and Kristy Reynolds, 2003). From consumers' perspectives, shipping charges are additional costs occurred in auctions. Therefore, when making a bidding decision, consumers have to take these delivery costs and risks into account. In Taiwan, self-pickup is a common alternative to shipping services, so we include this option into this study.

3. RESEARCH HYPOTHESES

The goal of this study is threefold. The first one is to compare how consumer evaluate different sellers based on judgment of bidding intention and seller trustworthiness. The second is to understand how consumer trade-off between four auctions attributes. The third one is to examine demographic and web

experience affect consumer trade-offs. Accordingly, we propose and will test the following hypotheses:

H1: Consumer’s preference for online auctions attributes will vary, depending on the type of evaluation (bidding intention and seller trustworthiness) being asked of the respondent

H2: In evaluated judgment of bidding intention, consumer’s preference for online auctions attributes will vary, depending on the demographic variables and web experience of the respondent

H3: In evaluated judgment of seller trustworthiness, consumer’s preference for online auctions attributes will vary, depending on the demographic variables and web experience of the respondent

4. RESEARCH METHOD

4.1 Subjects

There are 221 college students (122 men and 99 women), from two northeastern colleges in Taiwan, participated in this study, and their demographic variables and web experience variables were collected. The average age of the subjects was 21 years old, 44.3 percent of them had online auctions experiences.

4.2 Experimental design

A conjoint analysis was used to investigate subject’s bidding intention and trust for a variety of used textbook sellers that differed in terms of bidding price, seller rating (feedback score), payment type, and shipping charge. A used computer textbook with list price of NT\$800 was used as the product auctioned on the web. In this experiment design, subjects were asked to consider and make preference choices among hypothetical seller profiles, and decide on their bidding intentions and seller’s trustworthiness. In this study, each seller profile had four attributes, i.e. bidding price, seller rating, payment type, and shipping charge, and each attribute had three levels as shown below,

- (1) Bidding Price: a. NT\$480, b.NT\$400, c.NT\$320
- (2) Seller rating: a. 650, b.350, c.50
- (3) Payment types: Cash (or money orders), Bank transfer, Credit card
- (4) Shipping charges: a. Buyer pays shipping (NT\$80 dollars), b. free of charge, c. self-pickup

Because a full factorial design for this study would require 81 (3⁴) profiles, an orthogonal fractional factorial design (of 9 profiles) was used to reduce the number of attribute combinations and thus make the task of decision making more manageable.

The experiment scenarios addressed were similar to that encountered in full-profile rank conjoint analysis; subjects were asked to give a rank order of preference

for the 9 full profiles constructed by using the orthogonal main-effect design as Table 1.

Table 1 : Full profiles constructed by using orthogonal design

No.	Price	Seller rating	Payment types	Shipping charges
1	NT\$240	350	Credit card	Buyer pays
2	NT\$240	50	Cash	Free of charge
3	NT\$320	650	Credit card	Free of charge
4	NT\$320	50	Bank transfer	Buyer pays
5	NT\$320	350	Cash	Self-pickup
6	NT\$400	350	Bank transfer	Free of charge
7	NT\$400	650	Cash	Buyer pays
8	NT\$240	650	Bank transfer	Self-pickup
9	NT\$400	50	Credit card	Self-pickup

Figure 1 is an example of profile card, and every respondent was given 9 profile cards to sort out his or her ranking.

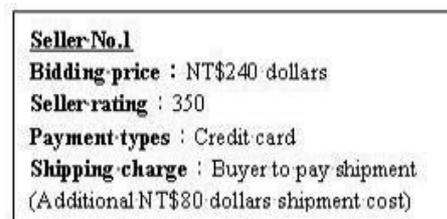


Figure 1 : Seller No.1 profiles card

5. RESULTS

5.1 Conjoint analysis

This research examined how various characteristics of auctions terms and the basis of consumers' evaluations affect their judgments about online auctions alternatives. The multivariate statistical method best suited for examining the trade-offs consumers are willing to make is a conjoint analysis. Conjoint analysis is a technique that requires subjects to make a series of judgments based on a set of attributes from which the underlying structure of their decisions can be decomposed. Conjoint analysis provides us with information about which attributes of an auction alternative are more important, and what trade-offs respondents are willing to make.

Table 2 shows the results of conjoint analysis. Bidding price is the most important attribute (28.60%) in respondent’s bidding intention judgment; the next important factor is seller rating (25.25%); while payment type (22.41%) and shipping charge (23.73%) are relatively less important. Utilities of each attribute

level are also shown in the table. Figure 2 indicates the utilities of each level for four attributes. The utility lines of bidding price shows that consumer prefers low price, and there is an increase in the marginal utility from NT\$400-NT\$320 to NT\$320-NT\$240 price interval. Seller rating has positive correlation with respondent's utility, in a diminishing marginal utility form. In addition, bank transfer and free of shipping are most preferred in attribute of payment type and shipping charges.

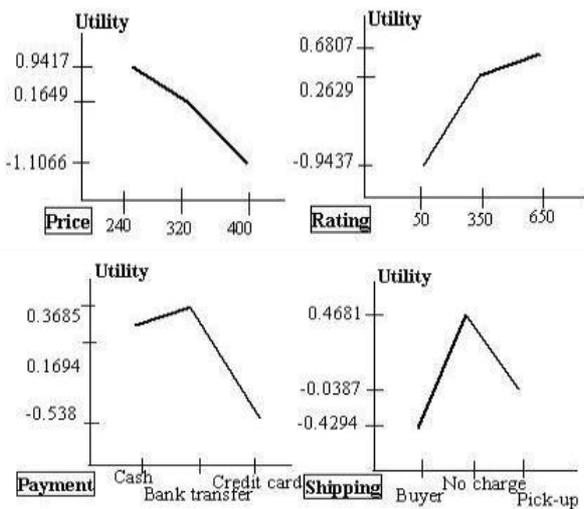


Figure 2 : Utility plot of four attributes in bidding intention evaluation

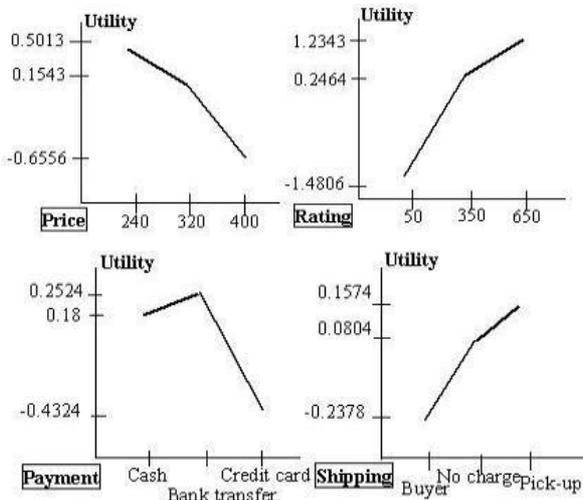


Figure 3 : Utility plot of four attributes in seller trustworthiness evaluation

When asked of seller trustworthiness, consumer weighs highly on seller rating (37.46%), while price (22.44%), payment type (20.26%) and shipping charges (19.84%) are far less important. Figure 3 shows the utility lines for each attributes. The only difference between bidding intention and seller trustworthiness evaluation is that respondents prefer self-pickup to other shipping charge options. The reason might be that self-pickup allows buyers to check the product before the deal is completed.

The results support hypothesis 1.

5.2 Test of H1

The data collected from respondent allow comparison of what online auctions attributes consumer desire when making judgment of bidding intention and seller trustworthiness. Since the two types of evaluations are within-subject data, we conduct a paired sample t test to examine whether the importance of each attribute is the same. The results show that the relative importance of bidding price ($t=3.497, p=.001$), seller rating ($t=-6.838, p=.000$), and shipping charges ($t=2.696, p=.008$) are significantly different, which means that when consumers consider of bidding for certain auctions, they focus on different auctions attributes than when they are asked to evaluate seller's trustworthiness. Table 3 shows that bidding price and shipping charge are weighed more when respondent intends to bid, while seller rating is significantly more important when respondent evaluate seller trustworthiness. Though seller rating might play an important role in consumer's bidding intentions, the results indicate that consumer's perceptions of these two concepts are quite distinct.

Table 2: Conjoint analysis of online auctions alternatives

Level	Bidding Intention		Seller Trustworthiness	
	Utilities	Importance	Utilities	Importance
NT\$400 NT\$320 NT\$240	-1.1066 0.1649 0.9417	28.60%	-0.6556 0.1543 0.5013	22.44%
650 350 50	0.6807 0.2629 -0.9437	25.25%	1.2343 0.2464 -1.4806	37.46%
Cash Bank transfer Credit card	0.1694 0.3685 -0.5380	22.41%	0.1800 0.2524 -0.4324	20.26%
Buyer pays No charge Self-pickup	-0.4294 0.4681 -0.0387	23.73%	-0.2378 0.0804 0.1574	19.84%

Table 3: Paired sample t test for evaluation types

Attribute	Mean differences	St. error measures	t	p (2-tailed)
Price	5.8424	1.6706	3.497	.001
Seller rating	-12.0619	1.7640	-6.838	.000
Payment type	2.0519	1.4289	1.436	.152
Shipping charges	3.6361	1.3489	2.696	.008

5.3 Test of H2

In this study, respondents' demographic variables (such as sex and age) as well as web experiences (i.e. Internet experience, online purposes, Internet usage, online auctions experience, and spending online) were collected. We use ANOVA to test how these variables affect the preferences of each attributes when making evaluation of bidding intention. The results of this analysis are explicated in Table 4.

5.3.1 Sex. There is significant effect of sex on the preference of shipping charges ($F=5.902$, $p=.016$). Female respondents concern more on shipping charges ($\bar{x}=26.44$) than male respondents ($\bar{x}=21.46$) when being asked of bidding intention.

5.3.2 Age. Age has significant effect on both bidding price ($F=12.514$, $p=.000$) and seller rating ($F=5.074$, $p=.007$). The data indicates that older respondents are more concerned of bidding price, and less concerned of seller rating. Further analysis shows younger respondents are less price sensitive than older ones.

Table 4: ANOVA for evaluation of bidding intention

Attributes ^a	BP	SR	PT	SC
Sex	0.012 ^b	0.296	2.965	5.902*
Age	12.514*	5.074*	1.865	1.773
Internet experience	1.682	2.413	0.326	0.597
Online purpose	0.259	0.279	1.224	1.594
Internet usage	1.014	0.853	0.843	0.739
Online auctions experience	10.838*	13.092*	0.154	0.108
Spending online	4.866*	0.471	3.344*	1.562

Notes: numbers in cells are statistic F

^a BP: Bidding price; SR: Seller rating; PT: Payment type; SC: Shipping charges

* Significant at $p<.05$

5.3.3 Internet experience, online purpose, and Internet usage. Preferences of online auctions attributes are not significantly influenced by Internet experience, online purpose, and Internet usage.

5.3.4 Online auctions experience. Bidding price is significantly more important to consumers with auctions experience online, and less important to ones without any online auctions experience ($F=10.060$, $p=.000$). On the other hand, respondents with auctions experience weigh less on seller rating than respondents without any experience ($F=13.849$, $p=.000$).

5.3.5 Spending online. Respondents' spending amounts online per month significantly affect their utilities of bidding price ($F=4.866$, $p=.001$) and payment type ($F=3.344$, $p=.013$).

5.4 Test of H3

The following analysis will use demographics and web experience as independent variables to test if the utilities

of each attributes are the same, when respondents are asked of seller trustworthiness evaluation. Table 5 shows that sex, web experience, and spending online do not have significant effect on auctions attributes. The effects of age, online purpose, and internet usage, and online auctions experience are reported below.

5.4.1 Age. There is significant effect on bidding price by age ($F=8.079$, $p=.000$). Younger respondents weigh more on bidding price for cues of seller's trustworthiness than older ones.

5.4.2 Online purpose. The purpose of online has significant effect on bidding price ($F=3.088$, $p=.028$). Consumers with online purposes of finding product information concern more on bidding price.

5.4.3 Internet usage and online auctions experience. Shipping charge is a significant important factor for intensive web users ($F=3.155$, $p=.025$) and experienced auctions consumers ($F=7.252$, $p=.008$).

Table 5: ANOVA for evaluation of seller trustworthiness

Attributes ^a	BP	SR	PT	SC
Sex	2.837	2.067	0.446	0.078
Age	8.079*	1.393	0.030	2.533
web experience	2.636	0.468	0.592	0.528
online purpose	3.088*	0.259	2.667	1.847
internet usage	1.012	1.232	0.476	3.155*
online auctions experience	2.639	0.635	1.248	7.941*
Spending online	2.393	0.478	0.218	0.710

Notes: numbers in cells are statistic F

^a BP: Bidding price; SR: Seller rating; PT: Payment type; SC: Shipping charges

* Significant at $p<.05$

6. CONCLUSION

This article provides an experimental test of consumer's trade-offs in online auctions context. These findings are essential because they help identify the importance of those attributes of online auctions under the consideration of bidding intentions and seller's trustworthiness. The results indicate significant differences between the types of evaluative judgment. The findings indicate that the way consumer perceives trustworthiness of a seller is quite different from their bidding intention.

This research examined how various characteristics of auctions terms and the basis of consumers' evaluations affect their judgments about online auctions alternatives. Using conjoint analysis and ANOVA, we were able to examine subjects' utilities for sellers' offerings that differed along four attributes as well as how these utilities were affected by demographic variables and web experience variables. The results indicate that price is the most important factor in bidding intentions (relative importance: price 28.6%; seller rating: 25.25%;

payment type: 22.41%; shipping charge: 23.73%), whereas the seller rating plays a more important role in consumer's trust in online auctions (relative importance: price 22.44%; seller rating: 37.46%; payment type: 20.26%; shipping charge: 19.84%). Payment type and shipping charge are also important factors, since utilities for these two factors are significantly greater than zero. The demographic variables show some influences on the utilities, for instance, female buyers are more sensitive to shipping charges, and older consumers put more weights on seller's ratings. As for web experience, the more experienced consumers are more low-price prone, and have significantly different judgments as compared to less experienced ones.

Interestingly, this paper suggests auctions sellers may succeed by simply compete on price. However, this conclusion could only be generalized in used textbook category; further research could include additional product categories as exemplars and different auctions attributes as factors. Although price and seller rating seem to be relatively more important factors, further research could examine factors in addition to payment type and shipping charge of the auctions. Of course, further testing should go beyond the college student sample.

REFERENCE

- [1] Ariely, D., Roth, A., & Ockenfels, A., "On-line Auctions: eBay and Amazon", (working paper). Cambridge, MA: M.I.T, 2000.
- [2] Bailey, J., *Electronic commerce: price and consumer issues for three products: books, compact discs, and software*. Organization for Economic Co-Operation and Development, Vol.4, 1998.
- [3] Benjamin, R., & Wigand, R., "Electronic markets and virtual value chains on the information superhighway", *Sloan Management Review*, Vol.36, No.2, pp62-72, 1995.
- [4] Brynjolfsson, E., & Smith, M. D. "Frictionless Commerce? A comparison of internet and conventional retailers", Working paper, 1999.
- [5] Choi, S. Y., Stahl, D. O., & Whinston, A. B. *The economics of electronic commerce*. Indianapolis: Macmillan, 1997.
- [6] Dholakia, U. M., Basuroy, S., & Soltysinski, K. "Auction or agent: A study of moderators of the herding bias in digital auctions", *International Journal of Marketing*. (in press).
- [7] Dholakia, U. M., & Soltysinski, K., "Coveted or overlooked? The psychology of bidding for comparable listings in digital auctions", *Marketing Letters*, Vol.12, pp223-235, 2001.
- [8] Eric B. Budish and Lissa N. Takeyama, "Buy price in online auctions: irrationality on the internet?" *Economics Letters*, Vol.72, pp 325-333, 2001.
- [9] Gilkeson, James H. and Kristy Reyonlds, "Determinants of Internet auction success and closing price: Aexploratory study", *Psychology & Marketing; ABI/INFORM Global*, Vol.20, No.6, pp537, 2003.
- [10] Green, P. E., & Srinivasan, V. "Conjoint analysis I n consumer research:Issues and outlook", *Journal of Consumer Research* ,Vol.5, pp105, 1978.
- [11] Guay, D., & Ettwein, J. , "Internet Commerce Basics", *Electronic Markets*, Vol.8, No.1, pp12-15, 1998.
- [12] Hair, Jr. J. F., et. al., *Multivariate Data Analysis with Readings* , 4th ed., N.Y.:Macmill-an, 1995.
- [13] Hanson, Ward, *Principles of Internet Marketing*. South-Western Publishing, Inc, 2000.
- [14] Klein, Stephan and O'keefe,Robert M., "The Impact of the Web on Actions: Some Empirical Evidence and Theoretical Considerations", *International Journal of ElectronicCommerce*, Vol.3, No.3, pp7-20, 1999.
- [15] Lee, H. G., "Do electronic marketplaces low the price of goods?" *Communications of the ACM*, Vol.41, No.12, 1997.
- [16] Luce, R. D. and J.W. Tukey, "Simultaneous Conjoint Measurement:A New Type of Fundamental Measurement", *Journal of Psychology*, Vol.1, pp1-27, 1964.
- [17] Smith, M. D., Bailey, J., and Brynjolfsson, E., *Understanding the Digital Economy*. Cambridge, MA: MIT Press, 1999.
- [18] Sulin Ba, Andrew B. Whinston, Han Zhang, "Building trust in online auction markets through an economic incentive mechanism", *Decision Support Systems*, Vol.35, pp273-286, 2003.
- [19] Zwass, V., "Electronic commerce: structure and issues", *International Journal of Electronic Commerce*, Vol.1, No.1, pp3-23, 1996.

Cross-Cultural Differences of Online Store Image Perception and Customer Purchasing Intention

Qinyu Liao, J. P. Shim

Dept. of Management and Information System
Mississippi State University, Mississippi State, MS, 39762, U.S.A.
ql15@msstate.edu¹, jshim@cobilan.msstate.edu²

ABSTRACT

In this study, seven factors affecting consumer perception of online store image are identified based on store image measurement of traditional stores. Enjoyment and trustworthiness are two new antecedents suggested for the online store image. A cultural dimension is also introduced to understand the possible differences of model structure and relationships between online store image perception, attitude toward purchasing at an online store and customer purchasing intention in a cross-cultural setting.

Keyword: Online store image, purchasing intention, cross-cultural difference

1. INTRODUCTION

Online shopping revenues and related corporate profits are still meager, though the industry is optimistic (Hoffman et al., 1999). Generating revenues using an online store is one of the key issues facing electronic commerce practitioners today.

Most studies of online shopping attitudes have attempted to identify factors influencing online shopping attitudes and behavior by taking different perspectives and focus on different factors in different ways. Store image is a multi-faceted construct that has been rigorously researched for "traditional stores". Several researchers have been able to successfully link "store image" to "intention to purchase" (Nevin and Houston, 1980; Grewal et al., 1998).

A significant number of online stores are designed and developed for firms with global reaching, but most of them address different cultural audiences with only linguistic difference (Chau et al., 2002). Studies have found cultural differences in using information systems (Nelson, et al., 1992; Ho et al., 1989). Understanding cultural and national effects on the use and management of IT is vital for multi-national firms.

The purpose of this study is to explore factors affecting consumer perceived online store image and its effect on subsequent purchasing intention, and proposed that online store image dimensions and their effect on attitude towards online store and purchasing intentions differ cross-culturally.

2. THEORETICAL MODEL

As both the presence and operation of online store depend heavily on IT, they are often regarded as a type of IS. Hence, when consumers accept and use

online store, they are accepting and using technologies and innovations.

Derived from TRA (Ajzen and Fishbein, 1980), TAM (Davis, 1989) has been proven successful in prediction and explaining behavior across a wide variety of domains including IT usage. The model hypothesizes that actual system use is determined by users' attitudes toward using. Attitude is directly affected by beliefs about the system, which consists of perceived usefulness and perceived ease of use. Researches have successfully adopted TAM to study the acceptance of Internet related technologies, such as email (Gefen, 1997), the World Wide Web (Fenech, 1998), and E-commerce (Chai, 2002). Therefore, using it as the basis for studying consumer perception of online store image is a valid approach.

Although TRA has been tested cross-culturally, it has been criticized for assuming western perspectives on time orientation, probabilistic thinking and locus of control (Cote and Tansuhaj, 1989). The model best captures the scientific, abstracting, context independent nature of western thought, coupled with its extensive empirical validation among western samples (Eagly and Chaiken, 1993). To model intention formation and prediction, cultural impact should be considered by using measures that reflect culturally-sensitive operationalizations of the underlying latent constructs (Malhotra and McCort, 2001).

3. CONCEPTUAL DEVELOPMENT

The online store is "a storefront in cyberspace, a place where customers can shop from their home computers and where merchants can offer merchandise and services for a fraction of the overhead required in a physical storefront" (Yesil, 1997).

Store image is the “personality” the store presents to the public. Studies have identified dimensions consumers perceive stores, which collectively make up store image (Hirschman, Greenberg and Robertson, 1978). Dickson and Albaum (1977) refined existing instrumentations for retail store image and ultimately arrived at the following dimensions: prices, products, store layout and facilities, service and personnel, promotion and “others.” Store image has been used as a predictor for numerous variables, including attractiveness of shopping area (Nevin and Houston, 1980) and purchase intention (Grewal et al., 1998). Creating or changing a shopping

area’s image is often a prerequisite to getting consumers to change their existing shopping intentions and behavior (Nevin and Houston, 1980).

The antecedents of an individual’s intention to purchase online store are likely to contain elements that are related to online store image. Jarvenpaa et al. have investigated empirically the relationship between intention to shop online, attitude towards shopping online and trust in the online store (Jarvenpaa, Tractinsky and Vital, 2000).

Online store image

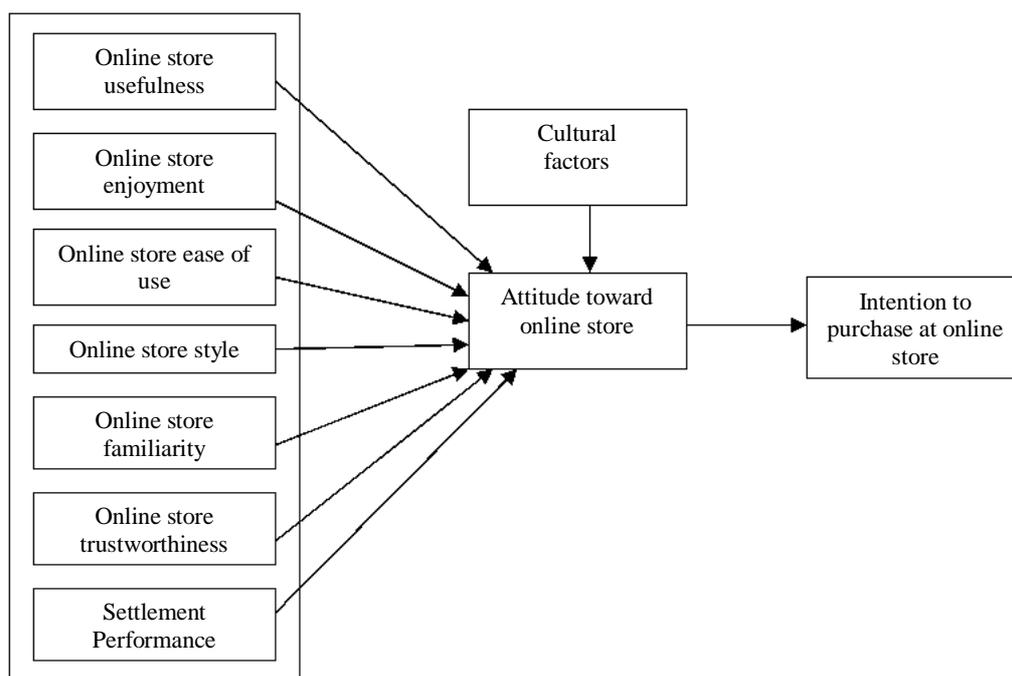


Figure 1. Research Model

Based on online store researches, seven factors that could affect consumer online store image perceptions are identified. They are online store usefulness, online store enjoyment, online store ease of use, online store style, online store familiarity, online store trustworthiness, and settlement performance. The research model in this study was developed based on TAM and these factors (Figure 1). A cultural dimension is added as the moderator.

3.1 Online Store Usefulness

Online store usefulness refers to the user’s subjective probability that using a specific online store will increase his or her performance in purchasing and information searching (Davis, 1986). The attributes include product variety, brands, price and product-retailing channel fit. The quality and usefulness of the information the customers found at the online store will

determine their prediction of satisfaction with the product prior to the actual purchase (Chen, 2001).

Proposition1: *The usefulness of an online store positively affects a consumer’s attitude toward purchasing at an online store.*

3.2 Online Store Ease of Use

Online store ease of use is the degree to which a user expects the use of the target online store to be free of effort (Davis, 1986). Some of the attitudes are easy-understanding, easy-finding, easy-learning, information focus (Lederer et al., 2000), and time taken for navigation and page download (Ranganathan and Ganapathy, 2002).

Studies have found out that consumer computer experience and use of other in-home shopping methods

influence a consumer's propensity to purchase at an online store.

Proposition 2: *The ease of use of an online store positively affects consumer's attitude toward purchasing at an online store.*

3.3 Online Store Style

The online stores face the heterogeneity of user profile. An online store that can attract, sustain and retain customers should have friendly interfaces, pleasant information presentation, and personalizable and supportive information services, such as multiple versions for different connectivity, chat room and email.

The usage of colors, multimedia, animation and graphics can reach customers on multiple cognitive levels and result in higher retention (Ranganathan and Ganapathy, 2002).

Ducoffe (1996) notes that irritating banner ads may be perceived as an unwanted, offending, and negative influence.

Proposition 3: *The online store style positively affects a consumer's attitude toward purchasing at an online store.*

3.4 Online Store Familiarity

Online store familiarity refers to customer's recognition of products, brand names and store names (Grewal et al., 1998). The store name is a form of the category-based processing perspective of store image (Keaveney and Hunt (1992). Brand name is a commonly used extrinsic cue to infer quality perceptions and can represent an aggregate of information about a product and its quality (Richardson, Dick and Jain, 1994; Dodds et al. (1991). Online shopping is closely related to recognition and recall (Chao, Au and Tam, 2000). Exposure to the brand name or store name gives customers a certain degree of familiarity.

Proposition 4: *The online store familiarity positively affects a customer's attitude toward purchasing at an online store.*

3.5 Online Store enjoyment

Enjoyment/playfulness refers to the pleasant and fun experience a consumer have with an online store, such as the eye-catching dynamic banners, enlivening background music or other multimedia online broadcasting, chat rooms, and online games. Enjoyment/playfulness has been suggested as an important dimension of consumer attitude towards online store (Liu et. al, 2000; Koufaris et al., 2002). Ducoffe (1995; 1996) evidences that enjoyment is positively related to online store advertisement value

and attitude to online advertising, and has been empirically linked to attitude towards using a website (Heijden et al., 2001; Webster, 1992).

Proposition 5: *The online store enjoyment positively affects a consumer's attitude toward purchasing at an online store.*

3.6 Online Store Trustworthiness

Trust in an online store is the subjective probability by which consumers expect that an online retailer will perform a given transaction in accordance with their confident expectations (Chai and Pavlou, 2002). Trust gives the consumer trust beliefs that they will gain control over a potentially uncertain transaction. Consumers have not widely adopted e-commerce (Hoffman et al., 1999), primarily because of trust issues (Palmer, Bailey, and Faraj, 2000) related to online retailers who have the chance to behave in an opportunistic manner by taking advantage of the distant and impersonal nature of online stores.

Consumers would be hesitant to transact with an online store that fails to signal a sense of trustworthiness. On the other hand, trust creates positive feelings towards transactions with online stores, providing expectations for a satisfactory feeling towards transactions with online stores, providing expectations for a satisfactory transaction. High-involvement goods, such as pianos, mortgages and intercontinental flight tickets, are likely to have a more severe impact on the relationship between trustworthiness and attitude to purchase (Heijden and Verhagen, 2002).

Jarvenpaa et al. (2000) discovered that higher consumer trust also reduces perceived risks associated with Internet shopping and generate more favorable attitudes towards shopping at a particular store, which in turn increases willingness to purchase from that store.

Proposition 6: *The online store trust-worthiness positively affects a consumer's attitude toward purchasing at an online store.*

3.7 Settlement Performance

This is a construct first introduced by Heijden and Verhagen (2002), which refers to the online retailers' handling of delivery, post-sales service, payment options, warranties and returns.

Proposition 7: *Online store settlement performance positively affects a consumer's attitude toward purchasing at an online store.*

3.8 Attitude Toward Online Store

Attitude is the overall evaluation of the desirability of a potential transaction with a specific online store.

Attitude has been proposed to influence behavioral intentions in multiple theories, such as the TPB (Ajzen, 1991) and the TRA (Fishbein and Ajzen, 1975). Applied to the online store consumer behavior context of this study, favorable attitude towards an online store is likely to encourage consumers to acquire information, to facilitate the act of providing information, and also to conduct monetary transactions with a given online retailer.

3.9 Intention to Purchase at Online Store

Intention to purchase at an online store is the user's likelihood to use an online store (Ajzen and Fishbein, 1980), maintain business relationships, and conduct business transactions. This is usually measured by consumers' willingness to buy and to return for additional purchases. Baker et al. (1994) found a link between store image and intention to purchase a product. Even if consumers finally decide not to purchase, it is important for marketers to know how to bring consumers to the online store, initiate communication, collect valuable feedback information, and encourage future transactions (Hoyer and MacInnis, 2001).

Proposition 8: A consumer's attitude toward purchasing at an online store positively affects his/her intention to purchasing at an online store.

3.10 Moderating Effect of Cultural Factors

Culture is a collective phenomenon. People learn patterns of thinking, feelings, and potential acting from living within a defined social environment, normally typified by country. Individual consumer tastes and purchasing patterns are thus partly determined by the collective values of their local community (Chau et al., 2002). If this notion is taken into the online store, different responses may be generated toward a globally generic online store design.

Empirical research has found significant cultural differences across international user communities and are further supported by Hofstede's work. He concluded that research in one culture may not be generalizable to another.

According to Hofstede's cultural consequences theory (1980), there are five dimensions in a culture: power distance, uncertainty avoidance, individualism versus collectivism, and masculinity versus femininity, and long-term versus short-term orientation. A cultural dimension is relevant to attitude toward online store.

Individualism/collectivism refers to the basic level of behavior regulation. People high on individualism view self and immediate family as relatively more important than the collective. Collectivist societies focus on maintaining harmony by going along with the in-group's wishes and promoting long-term relationships (Bond and Smith, 1996). In general, once collectivist

societies establish a positive attitude toward something, they tend to internalize it and take it into their in-group circle. Thus, we would expect that a collectivist culture would influence its members to want to maintain harmonious relations with those vendors. Members of individualist cultures prefer self-sufficiency and should feel no mutual influence toward the maintenance of group harmony. Chau et al. (2002) suggested that online testimonials and links to related chat groups could strengthen the social communications of the collectivist consumers who respect social relationships and their shared loyalty.

Short versus long-term orientation is people's basic reference time frame; short-term involves the tendency toward consumption and maintaining materialistic status, and long-term suggests thrift, perseverance, following tradition, and deferred satisfaction. According to Hofstede (2001), a culture of long-term orientation emphasizes persistence and thrift, in addition to respect for tradition and people feel free to put off making purchase at an online store until they are comfortable with it. Consumers of a long-term orientation society may also be less demanding of immediately online store satisfaction and have higher tolerance of errors or misdelivery if they can see a trend toward incremental quality improvement. For a short-term oriented consumer, the time concept is absolute and demand for temporal precision is high.

When masculinity is high, online store consumers may act more aggressively in their intentions and expression. This action may be more self-centered and less group-oriented.

The uncertainty avoidance describes the extent to which a society feels threatened by uncertain and unstructured situations and the eagerness to avoid such situations. This orientation may emphasize controls achieved in strong procedural and technically oriented online store navigation. A less structured online store or unfamiliar store name could prevent the consumers from using it.

Power distance is the degree to which the less powerful accept the prevailing distribution of power, and the extent of adherence to formal authority channels. High power distance cultures have members who are much more comfortable with centralized power than lower power distance cultures. For an online store, customers in a low power distance society would be less desirable to influence or control the services they received.

Proposition 9: Online store image dimensions and their effect on attitude toward online store and purchasing intentions differ cross-culturally.

4. SUMMARY AND RESEARCH IMPLICATIONS

One of the most basic concerns of online stores is

finding and retaining customers. Our work tapped on this question from the online store image determinants perspective and identified seven antecedents of online store image and their influence on the intention to purchase online. They are the perceived online store usefulness, perceived online store ease of use, online store enjoyment, online store style, online store familiarity, online store trustworthiness, and settlement performance. The most effective way for online retailers to develop profitable exchange relationship with online consumers is to earn their trust by have a trustworthy online store image. As more and more competitors are having similar online presence which convenience nearly costless switching for online shoppers, the enjoyment of an online store becomes an important consideration to sustain the interest of shoppers and earn their loyalty over other online stores.

The inclusion of the cultural dimension help augment the explanation of online store acceptance and usage by consumers across culture, which is very necessary as more and more online stores targeted at global customers.

Our findings can help online retailers predict consumer's willingness to accept and use their online stores, diagnose the reasons for low acceptance for existing stores, and develop corrective steps. The central point is consumer attitude towards using online stores. If consumers have more positive attitude, they are more likely to accept and use online stores. The model also provides insights on how to achieve this: a more positive consumer attitude can be achieved by improving an online store's image.

While the patterns of online consumer purchasing are gradually becoming standardized, localized online store addressing local consumer preferences are also increasing (Chau, 2002). Therefore, retailers should tailor the online store by taking cultural differences into consideration.

A number of limitations are inherent to this study. We will only study the influence of online store image and cultural factor moderation on intention of purchasing at an online store which is different from actual purchase. There are more factors needs to be explored as how the mental process of making a purchase decision made. In this study, we use cultural influence as a collective factor, future researches should explore the possible interactions between the dimensions of culture for more insightful conclusions.

REFERENCES

- [1] Ajzen, I. "The theory of planned behavior", *Organizational Behavior and Human Decision Processes*, Vol. 50, pp. 179-211, 1991.
- [2] Ajzen, I., Fishbein, M. *Understanding Attitude and Predicting Social Behavior*. Prentice Hall, Englewood Cliffs, NJ, 1980.
- [3] Baker, J., Grewal, D., and Parasuraman, A. "The influence of store environment on quality inferences and store image", *Journal of the Academy of Marketing Science*, Vol. 22, No. 4, pp. 328-339, 1994.
- [4] Bond, R., and Smith, P. B. "Cross-cultural social and organizational psychology", *Annual Review of Psychology*, Vol. 47, pp. 205-235, 1996.
- [5] Chau, Y. K., Au, G., and Tam, K. Y. "Impact of information presentation modes on online shopping: an empirical evaluation of a broadband interactive shopping service", *Journal of Organizational Computing and Electronic Commerce*, Vol. 10, No. 1, pp. 1-22, 2000.
- [6] Chai, L., Pavlou, P. A. "Customer relationship management.com: a cross-cultural empirical investigation of electronic commerce", *Proceedings of 8th Americas Conference on Information System*, Dallas, TX, 2002.
- [7] Chau, P. Y. K., Cole, M. Massey, A. P., Montoya-Weiss, M. and O'Keefe, R. M. "Cultural Differences in the Online Behavior of Consumers", *Communications of the ACM*, Vol. 45, No. 10, pp. 138-143, 2002.
- [8] Chen, L. D., Gillenson, M. L., and Sherrell, D. L. "Enticing online consumers: an extended technology acceptance perspective", *Information & Management*, Vol. 39, pp. 705-719, 2001.
- [9] Cote, J. A. and Tansuha, P. S. "Culture bound assumptions in behavior intention models", *Advances in Consumer Research*, Vol. 16, pp. 105-109, 1989.
- [10] Davis, F. D. "A technology acceptance model for empirically testing new end-user information systems: theory and results", *Doctoral Dissertation, MIT Sloan School of Management*, Cambridge, MA, 1986.
- [11] Davis, F. D. "Perceived usefulness perceived ease of use and user acceptance of information technology", *MIS Quarterly*, Vol. 13, No. 3, pp. 319-340, 1989.
- [12] Dickson, J., and Albaum, G. "A method for developing tailormade semantic differentials for specific marketing content areas", *Journal of Marketing Research*, Vol. 14, No. 2, pp. 87-91, 1977.
- [13] Dodds, W. B., Kent, B. M., and Grewal, D. "Effects of price, brand and store information on buyers' product evaluations", *Journal of Marketing Research*, Vol. 28, pp. 307-319, 1991.
- [14] Ducoffe, R. H. "How consumers assess the value of advertising", *Journal of Current Issues in Advertising*, Vol. 17, No. 1, pp. 1-18, 1995.
- [15] Ducoffe, R. H. "Advertising value and advertising on the web", *Journal of Advertising Research*, Sept./Oct., pp. 21-35, 1996.

- [16] Eagly, A. A. and Chaiken, S. *The Psychology of Attitude*, Harcourt Brace Jovanovich, Forth Worth, TX, 1993.
- [17] Fenech, T. "Using perceived ease of use and perceived usefulness to predict acceptance of the world wide web", *Computer Networks & ISDN Systems*, Vol. 21, No. 4, pp. 629-630, 1998.
- [18] Fishbein, M. and Ajzen, I. *Belief, Attitude, Intention and Behavior: an Introduction to Theory and Research*, Addison-Wesley, Rading, MA, 1975.
- [19] Gefen, D. and Straub, D. W. "Gender differences in the perception and use of e-mail: an extension to the technology acceptance model", *MIS Quarterly*, Vol. 21, No. 4, pp. 389-400, 1997.
- [20] Grewal, D., Krishnan, R., Baker, J., and Borin, N. "The effect of store name, brand name and price discounts on consumers' evaluation and purchase intentions", *Journal of Retailing*, Vol. 74, No. 3, pp. 331-352, 1998.
- [21] Heijden, H. V. D. and Verhagen, T. "Measuring and assessing online store image: a study of two online bookshops in the Benlux", *Proceedings of the 35th Hawaii International Conference on System Sciences*, 2002.
- [22] Heijden, H.V. D. "Factors influencing the usage of websites: the case of generic portal in the Netherlands", *Proceedings of 14th Bled Electronic Commerce Conference, Bled*, 2001.
- [23] Hirschman, E. "Store image and perception of performance in retailing", *Journal of Business Research*, Vol. 17, pp. 91-100, 1981.
- [24] Ho, T. H., Raman, K. S., and Watson, R. T. "Group decision support systems: the cultural factors", *Proceedings of the 10th International Conference on Information Systems*, Boston, MA, pp. 119-129, 1989.
- [25] Hoffman, D. L., Novak, T. P., and Peralta, M. "Building consumer trust online", *Communications of the ACM*, Vol. 42, No. 4, pp. 80-85, 1999.
- [26] Hofstede, G. *Culture's Consequences*, 2th Edition, Thousand Oaks: Sage, 2001.
- [27] Hofstede, G. *Culture's Consequences*, Beverly Hills: Sage, 1980.
- [28] Hofstede, G. and Bond, M. H. "The confucius connection: from cultural roots to economic growth", *Organizational Dynamics*, Vol. 16, No. 4, pp. 4-21, 1988.
- [29] Hoyer, W. D., and Macinni, D. J. *Consumer Behavior*, 2nd Edition, Boston: Houghton-Mifflin, 2001.
- [30] Jarvenpaa, S. L., Tractinsky, N., and Vitale, M. "Consumer trust in an internet store", *Information Technology and Management*, Vol. 1, pp. 45-71, 2000.
- [31] Keaveney, S. M., and Kenneth, A. H. "Conceptualization and operationalization of retail store image: a case of rival middle-level theories", *Journal of the Academy of Marketing Science*, Vol. 20, pp. 165-175, 1992.
- [32] Koufaris, M., Kambil, A., and LaBarbera, P. A. "Consumer behavior in web-based commerce: an empirical study", *International Journal of Electronic Commerce*, Vol. 6, No. 2, pp. 115-138, 2002.
- [33] Lederer, A. L., Maupin, D. J., Sena, M. P., and Zhang, Y. "The technology acceptance model and the world wide web", *Decision Support Systems*, Vol. 29, pp. 269-282, 2000.
- [34] Lindquist, J. D. "Meaning of image: a survey of empirical and hypothetical evidence", *Journal of Retailing*, Vol. 50, No. 4, pp. 29-38, 1974.
- [35] Liu, C. and Arnett, C. P. "Exploring the factors associated with web site success in the context of electronic commerce", *Information & Management*, Vol. 38, No. 1, pp. 23-33, 2000.
- [36] Malhotra, N. K. and McCort, J. D. "A cross-cultural comparison of behavioral intention models: theoretical considerations and an empirical investigation", *International Marketing Review*, Vol. 18, No. 3, pp. 235-269, 2001.
- [37] Nevin, J. R. and Houston, M. J. "Image as a component of attraction to intraurban shopping areas", *Journal of Retailing*, Vol. 56, No. 1, pp. 77-93, 1980.
- [38] Nelson, R., Weiss, L. R., and Yamazaki, K. "Information resource management within multinational corporation", *International Information Systems*, Vol. 1, No. 4, pp. 57-88, 1992.
- [39] Niederman, L. D., "Computer usage by small and medium sized European firms: an empirical study", *Information & Management*, Vol. 2, pp. 66-77, 1979.
- [40] Palmer, J. W., Bailey, J. P., and Faraj, S. 2000. "The role of intermediaries in the development of trust on the WWW: the use and prominence of trusted third parties and privacy statements", *Journal of Computer-Mediated Communication*, Vol. 5, No. 3, 2000.
- [41] Ranganathan, D., Ganapthy, S. "Key dimensions of business-to-consumer websites", *Information & Management*, Vol. 39, pp. 457-465, 2002.
- [42] Richardson, P. S., Alan, S. D., and Arun, K. J. "Extrinsic and intrinsic cue effects on perception of store brand quality", *Journal of Marketing*, Vol. 58, pp. 28-30, 1994.
- [43] Webster, J. and Martocchio, J. J. "Microcomputer playfulness: development of a measurement with workplace implications", *MIS Quarterly*, Vol. 16, No. 2, pp. 201-226, 1992.
- [44] Yesil, M. *Creating the Virtual Store: Taking Your Website from Browsing to Buying*, Wiley, New York, 1997.

E-Business Adoption in Marketing and its Relationship with Factors in SWOT Analysis: an Empirical Investigation of Small Software Businesses

Edward W.N. Bernroider ¹, Alexander Hampel ²

¹ Department of Information Business, Vienna University of Economics and B.A., Austria
Edward.Bernroider@wu-wien.ac.at

² Department of Computer Science and Business Informatics, University of Vienna, Austria
Alexander.Hampel@univie.ac.at

ABSTRACT

This research gives empirically grounded insights into e-business adoption in conjunction with factors usually considered in SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis in Austrian micro and small to medium sized software enterprises. The study evaluates the degree in which the Internet is used in supporting the firm's primary activities in marketing and sales based on primary data collected from 141 Austrian software firms. The results show that internet usage in this area is dependent on the size of the company. While almost all small to medium sized enterprises apply e-business to some degree, many micro enterprises do not seek to utilise possible potentials. The study reveals that firms perceiving their relative strengths on branding, pricing, product diversity, internationalisation, and access to new technologies have adopted more advanced e-business support. From the external perspective, organisations more intensively relying on e-business tend to experience less competitive pressure, have lower capital resources, and consider international markets more often as business opportunities than their competitors.

Keywords: software industry, e-business, e-marketing, SMEs, success factors

1. INTRODUCTION

Small businesses worldwide occupy an important economic position since they constitute key drivers of innovation, economic stabilisation, and employment [1-4]. This also seems to be the case for the Austrian computer software industry [5]. The software business itself is highly competitive, complex, knowledge-intensive, and characterised by a highly globalised as well as dynamic business environment. Software enterprises are therefore increasingly reliant on utilising new strategies to remain competitive. The Internet provides managers with opportunities to fortify and strengthen their competitive advantages [6, 7]. Recently published academic findings show a wide spectrum of possible effects of e-business on organisational performance [8]. The potential of e-business is believed to be a precondition for the success of future business organisations [6]. E-business is seen as a means to strengthen the competitiveness of firms [9]. Small software businesses tend to offer selective products and services connected with limited buying mobility of potential customers, instead of mass products where infinite multiplication of choices exist. It is believed that selective brands drive e-business as an strategic opportunity [10].

Empirical research that uncovers the impacts of e-business and relationships between utilisation rates and strengths, weaknesses, opportunities and threats (SWOTs) is underdeveloped. This paper answers the call for more exploratory, empirically grounded work [11] by focusing on: (i) e-business utilization, (ii) and interdependencies between SWOTs and e-business utilization. In all steps of analysis the focus lies on

micro as well as small to medium sized enterprises in the Austrian software sector. This article is restricted to Business to Consumer (B2C) e-business covering all offline and online marketing transactions using Internet based information and communication technologies. Marketing/sales is seen as the most frequently area impacted by e-business, i.e. it affects first ways in which the firm operates with the market [4]. The work assesses the impact of the Internet's growing influence on marketing strategy and possible relationships with perceived SWOTs. It does not elaborate on e-business models, frameworks, and adoption as well as operation practices.

The article is organised as follows. The chosen methodology and details concerning the acquisition of data are given in the next section. Thereafter, the frameworks for analysing the endogenous and exogenous situation are presented. This is followed by the analysis of the empirical data, which comprises the firm size distribution and e-business adoption in the industry as well as endogenous and exogenous perspectives. The last section concludes the article.

2. METHODOLOGY

The methodology employed is an empirical study of the Austrian computer software industry undertaken in the year 2000. The supplied address material for the study contained not only all Austrian software companies but also many hardware manufacturers and other IT firms. Therefore the empirical work undertaken had to consist of two steps. The first step allowed the screening of the provided address material in order to filter out the relevant software companies for the second main step of

analysis. In order to be included in the main sample, firms had to belong to the software industry, which was divided in three segments: Packaged software (SIC 7372, respectively NAICS 51121, 334611), custom software and software related services (SIC 7371, 7373 respectively NAICS 541511, 541512) [12]. Although custom software, i.e. software that is produced only once, is viewed as service output [13] this segment was considered separately. For both, the pre-liminary and the main analysis, the design of a questionnaire which was validated in several pre-tests was necessary. For the pre-liminary phase random sampling was employed and the 600 chosen companies were contacted by telephone. The achieved response rate was 70.4%. Some companies could not be contacted, because they had ceased to exist, the address was wrong or could not be found, etc. These neutral dropouts were considered in the calculation of the response rate and therefore did not decrease the return quota. On completion of the preliminary analysis structured face-to-face interviews followed in the main step of analysis, which based on a stratified and disproportional sample with subgroups according to company size. The rate of return for this second and main step of the study was 55.6%. After completing both, the pre-liminary and main step of analysis, non-response effects were examined. The analysis (regarding known distributions of variables of the Austrian computer software market such as the type of the organisation) revealed no significantly different characteristics between non-respondents and respondents.

The main questionnaire assessed general company information, the endogenous situation, exogenous data, and questions on co-operations. This endogenous section also covered the level of support of marketing objectives through use of Internet technology. The question was stated so that all organisations regardless of their software sub sector or their different e-business tactics were able to respond on their Internet utilisation rate. This included mentioning examples for online and offline e-business tools such as e-newsletters, web banners, web sites or portals. It is important to mention, that this article does not consider the area of e-procurement or e-sourcing. The assessed variable only refers to primary activities in marketing products and services.

3. SWOT ANALYSIS

Although there is an extensive amount of literature covering the strategic planning process, most approaches include a cyclic iteration of the following five elements. The strategic planning process begins with a statement of the corporate mission and goals (step 1). This is followed by the analysis of the organisation's external competitive environment (step 2) and the analysis of the organisation's internal operating environment (step 3). The internal analysis is followed by the selection (step 4) and the implementation of

strategies (step 5). The last step also involves the design of the organisational structure and control systems necessary to implement the chosen strategy [14, 15]. The focus of this article lies upon step 2 (external analysis) and 3 (internal analysis). The purpose of the external analysis is to identify opportunities and threats in the organisation's operating environment, while the internal analysis seeks to pinpoint the organisation's strengths and weaknesses. Many companies are conducting a SWOT analysis as part of the strategic planning process to identify the strengths, weaknesses, opportunities and threats before proceeding to the formulation of a strategy [16, 17]. SWOT analysis, respectively the analysis of "key" or "critical" success factors, belongs to the highest ranked set of techniques of strategic analysis used by firms in empirical surveys [18].

3.1 Internal Analysis

This work seeks to analyse if the endogenous situation faced by the firm is interrelated with the applied level of e-business initiatives. For the development of the framework this research builds on the generic building blocks of competitive advantage together with a resources and capability based view: Efficiency, quality, customer responsiveness and innovation are the generic building blocks of competitive advantage that every software company can adopt. Achieving superiority requires an organisation to develop appropriate competence, which in turn is a product of the kind of resources and capabilities that a company possesses [15]. The variables that influence the long-term business success were therefore categorised into six areas:

- Resources
- Capabilities
- Efficiency
- Quality
- Customer responsiveness
- Innovation

The distinctive competence of a software company arises from two complementary sources: its resources and capabilities. The distinction between resources and capabilities is important in understanding what generates specific competence. Without the appropriate resources, the company may not be able to create a distinctive competency [15]. The inquired variables contributing to the resource-based view were "acquisition of venture capital", "equity capital situation", "trademark/label" and "image". Variables used to assess the capabilities of the organisation were "internationalisation capabilities", "order/project procurement", "organisational structure", "adherence to costs and schedules", "employee motivation" and "employee qualification". While only efficiency was not divided into sub-areas, quality was analysed by the factors "product quality" and "business process quality", customer responsiveness was represented by the variables "product diversity", "price levels", "customer

satisfaction", "level of specialisation" and "customer orientation", and to innovation the variables "know-how embedded in company" and "access to new technologies" were attributed. Although the inquired variables were allocated into different groups, they are highly interrelated. For example, factors contributing to innovation such as "know-how embedded in the company" or "access to new technologies" can also enhance quality, efficiency and customer responsiveness.

3.2 External Analysis

The independent variables of the external perspective were subdivided into potential exogenous growth barriers (threats) and potential exogenous growth drivers (opportunities). The selection of variables for both groups was influenced by the Porter's five forces model [19] and a wider macro-environmental view. Industries are embedded in a wider macro-environment, which comprises the technological, social and demographic, ecological, political and legal, and broader macro-economic environment [15, 20]. Changes in the macro-environment can have a direct impact on the five forces in Porter's model.

The potential exogenous growth barriers deemed as relevant for the analysis in conjunction with e-business were:

- Marketing/distribution in foreign countries
- Culture, varieties in language
- Trade/commerce law issues
- Governmental regulations
- Lack of ITC standards
- Acquisition of capital
- Labour legislation issues
- E-business maturity in Austria
- Austrian market size
- TC costs and infrastructure in Austria
- Software piracy
- Rivalry

As potential exogenous growth drivers the following criteria were selected:

- EU Software patent protection initiative
- Break-up of the former eastern bloc
- EU east expansion
- EU domestic market
- Liberalisation of telecommunications

4. EMPIRICAL RESULTS

4.1 Distribution of company size

The empirical data showed a strong overbalance of smaller software enterprises in Austria: Micro-enterprises (ME) account for 55.7% of the Austrian software organisations, 32.2% can be attributed to small-to-medium sized enterprises (SME), and 12.1% to large enterprises (LE). Classification was performed

using data on number of employees, turnover and independency following the definition proposed by the Commission of the European Community (EC, 1996). The applied criteria and the approximated overall number of software enterprises in Austria are presented in Table 1. In Austria the average firm size has declined in line with smaller firms entering the industry. Especially at the end of the 1980's and the early 1990's annual growth rates of over 20% (measured by number of enterprises in Austria) caused the decline of average firm size.

Table 1. Company size distribution

Size	Emplo yees	Turn over Mio.€	Indep end. ¹	No. of comp. (rel.)	No. of comp. (abs.)
ME	1-9	< 7	< 25%	55,7%	1.082
SME	10-249	< 40	< 25%	32,2%	625
LE	> 250	> 40		12,1%	235
Total				100%	1.941

¹ Capital share in external ownership

In the following sections analysis was performed only for the MEs and SMEs which constitute the research focus of this article. Consequently, from the 174 datasets 91 MEs (weighted N=97) and 50 SMEs (weighted N = 56) were included in the analysis, whereas two SMEs did not respond on e-Business utilisation.

4.2 E-Business Adoption

This study assessed the degree in which the Internet is used in supporting the firm's primary activities in the marketing area on an interval-scale from 1 (never) to 5 (very intensively). To test whether MEs and SMEs have identical distributions the Mann-Whitney U Test was applied, which is a nonparametric alternative to the independent-sample Student t test. It revealed that significant differences ($p < 0.05$) exist. The data showed that in the mean SMEs are more intensively engaged in e-business compared with MEs. The following exploratory analysis was undertaken for each group of companies classified by their size separately.

The data set comprising only MEs exhibits a clear kurtosis and low skewness, i.e. indicating a very symmetric distribution. The variability of the data set is high. A great part of the organisations ignore e-business while approximately the same proportion of respondents already embraces e-business based transactions. Traditional marketing and communication channels are still predominant in the case of MEs. Many companies do not even build on the simplest e-based methods. Among SMEs, the distribution changes significantly. In SMEs, the utilisation rate of e-business is concentrated around a high level of utilisation. Only a small minority of SMEs have not implemented any e-business initiatives.

4.3 Internal Analysis

The endogenous success factors given in a previous section had to be rated by managers of Austrian software companies on an interval scale based on our local school grading system from 1 (very positive) to 5 (very negative) as perceived in their own organisation. The first step of analysis was to test if there is a relationship between e-business utilisation and endogenous success factors. As a non-parametric test of statistical significance for bivariate analysis, the Chi square test was applied followed by an analysis of cross tabs with a special focus on standardised residuals to identify influential observations [21]. The strength of a possible relationship between success factors and the size of the company was analysed by calculating the Spearman rank correlation coefficient. Again, a non-parametric measure was used, which does not make any assumptions about the underlying frequency distributions.

The following analysis was applied in multiple steps. The first step comprised MEs and SMEs grouped together, followed by separate steps of analysis for each group of companies. Viewing both MEs and SMEs together, the study identified seven out of the twenty success factors analysed that showed differences in how they were rated by managers dependent on the rate of e-business adoption. Among MEs, six variables remained. Considering SMEs, four factors showed a significant relationship (see Table 2). It can be noted, that due to the impact of sample size on statistical power, less statistical significance is achieved with SMEs, where the number of observations is lower (50) compared to MEs (91). The endogenous situation improves with e-business utilisation for almost all identified relationships. The analysis of residuals for non-correlated factors has shown the same tendency.

The internationalisation capabilities improve, especially in SMEs, with higher e-business adoption levels. This reflects that the use of information and communication technologies in general, decreases the costs of distance, especially for more product rather than service oriented firms. The Internet is an enabling technology to reach new customers and suppliers in an affordable, more efficient and effective way [9] without the necessity to set up and maintain local branches in foreign markets. For MEs, the data shows that firms that see their relative strengths in the areas of trademarks/labels, product diversity, and price levels are more intensively engaged in e-business. These factors seem to be key strengths that motivate software firms to market their products and services in the Internet, consequently not fearing the associated global transparency. The ability to access new technologies is positively related with higher e-business adoption levels. Another study reported that the highest barrier for SMEs to doing business on the internet is the lack of technical skills [22]. This research adds empirical evidence to this statement applicable to

the software industry. The empirical evidence supporting this relationship is limited to the sample comprising all companies. The significance was lost when analysing both groups defined by company size separately. Efficiency was expected to be better in organisations with more e-business supported operations. A significant relationship was detected for all companies analysed together, but no sensible pattern underlying the relationship was identified. The same situation applies to product quality. A surprising indication is that MEs with lower customer orientation levels have more widely adopted e-business. The link between firms and their customers is believed to be strengthened by enabling e-business. Finally, the ability to acquire venture capital improves with the level of e-business diffusion. In general, MEs and SMEs with venture capital background passed a sales and marketing due diligence, which gives an explanation for their above average inclusion of state-of-the-art information and communication technology in their marketing models.

Table 2. Identified endogenous success factors showing dependencies by Chi²-Test and/or Spearman Rank Correlation Coefficients

No.	Variable	Classification
1	Internationalisation	Capabilities
2	Product diversity	Customer responsiveness
3	Price levels	Customer responsiveness
4	Customer orientation	Customer responsiveness
5	Efficiency	Efficiency
6	Access to new technologies	Innovation
7	Product quality	Quality
8	Trademark/label	Resources
9	Acquisition of venture capital	Resources

To identify how and if the significant endogenous criteria relate as sets to the e-business utilisation variable, two stepwise regression analyses were performed for each group defined by company size. It was therefore implicitly hypothesised that a multivariate model can be found, that significantly predicts e-business utilisation as the dependent criterion. The results of the stepwise analyses are reported in Table 3. For MEs, five steps were calculated, i.e. while five of six factors retain in the regression equation, the trademark/label factor was omitted. Due to the stepwise approach, all remaining factors significantly contribute to the model and therefore explain e-business utilisation. The extent of change in e-business utilisation for each 1-unit change of predictors seems to be similar. The direction of the change is different for product quality and customer orientation compared with the other factors significantly contributing to the final model. Product diversity was the first measure to be included in the stepwise calculation. Specifically, 31% of the

variance in e-business utilisation is explained by the five success factors in the final model ($p < 0.01$).

Table 3. Results of stepwise regression analysis of e-business utilisation on significant endogenous factors in MEs

Step	Variable	B	R ²	corr. R ²	Δr^2
1	Product diversity**	-0.50	0.11	0.10	-
2	Product diversity**	-0.58	0.18	0.16	0.06
	Customer orientation**	0.56			
3	Product diversity**	-0.50	0.25	0.22	0.06
	Customer orientation**	0.65			
	Price levels**	-0.56			
4	Product diversity**	-0.46	0.31	0.28	0.06
	Customer orientation**	0.73			
	Price levels**	-0.55			
	Acquisition of venture capital**	-0.30			
5	Product diversity**	-0.50	0.35	0.31	0.03
	Customer orientation**	0.67			
	Price levels**	-0.57			
	Acquisition of venture capital**	-0.34			
	Product quality*	0.46			

* $p < 0.05$, ** $p < 0.01$

The regression constant was excluded from the table for every step.

For SMEs, four factors were included in the calculation, whereas two retained in the regression equation and together significantly explain e-business utilisation (see Table 4). With the final regression function, 16% of the variance in e-business utilisation is explained by the two variables ($p < 0.01$).

Table 4. Results of stepwise regression analysis of e-business utilisation on significant endogenous factors in SMEs

Step	Variable	B	R ²	corr. R ²	Δr^2
1	Internationalisation*	-0.24	0.10	0.09	-
2	Internationalisation**	-0.28	0.19	0.16	0.07
	Trademark/label*	0.45			

* $p < 0.05$, ** $p < 0.01$

The regression constant was excluded from the table for every step.

4.4 External Analysis

During the interview, the companies assessed the given exogenous variables according to their perceived influence on business growth on two different interval scales for each of the following groups: potential exogenous growth barriers and potential exogenous growth drivers. With these variables the same analysis as presented above was conducted. The analysis showed that observed e-business utilisation had a significantly

different distribution depending on a number of external factors more specifically in SMEs. An interesting indication is that managers in SMEs who are more intensively relying on e-business are experiencing less rivalry than their competitors. The marketing opportunity in the European Union is considered as a more important growth driver by software managers employed in firms with higher e-business utilisation. Finally, firms with lower capital resources more intensively exploit e-business. For the remaining factors, no clear direction of the dependencies was identified.

Table 5. Identified exogenous factors showing dependencies by Chi²-Test and Spearman Rank Correlation Coefficients

No.	Variable
1	EU domestic market
2	Labour legislation issues
3	Marketing/distribution in foreign countries
4	Acquisition of capital
5	Austrian Telecommunication costs and infrastructure
6	Rivalry
7	EU Software patent protection initiative

In analogy to the previous section, stepwise regression analyses were performed to complement the analysis with a multivariate method considering all detected significant variables together. Since no multivariate dependency profile was detected for MEs, the analysis was undertaken for SMEs only. The calculation process based on the identified exogenous factors as predictors was not able to conclude with a sensible and statistically significant model.

5. CONCLUSIONS

The paper provides empirically grounded insights into e-business utilisation in conjunction with related endogenous as well exogenous factors for MEs and SMEs in the Austrian software industry. In MEs, the usage of e-business tools in marketing their products and services is diverse with many organisations ignoring e-potentials. Among SMEs, e-business is variability is much less, i.e. most companies already seek to complement and enhance their marketing skills with e-business tools. They exploit e-business as an additional marketing channel, allowing for a global reach. They therefore follow Porter's argument that organisations should use the Internet in conjunction with their traditional business models and activities [7].

The analysis of the internal situation faced by the firm showed that a number of success factors were perceived as relative strengths in companies with better e-business support. E-business can be advantageous or disadvantageous to the business strategy of small firms [22]. Behind this background it does not seem surprising, that branding and pricing together with an attractive product portfolio were identified as key

strengths connected with the decision to engage the global market via the Internet. The ability to access new technologies seems to improve with higher levels of e-business adoption. If e-business is considered as a technology based innovation meant to redesign or complement primary activities in marketing and sales, it seems reasonable to expect that technology innovators among SMEs and MEs should show a higher e-business adoption rate. A surprising result was that customer orientation decreases with the level of e-business support in MEs. The external perspective embracing a number of factors regarded as threats and opportunities showed a less number of factors with significant relationships. Companies more intensively relying on e-business tend to experience less competitive pressure, have lower capital resources, and consider the international markets in the European Union as a more valuable business opportunity than their competitors.

In addition to correlation and dependency tests, multivariate regression analysis was applied to see if the variation of e-business utilisation can be explained by a multivariate consisting of the endogenous, respectively exogenous criteria for MEs and SMEs separately. The results showed that the multivariate relationships were found in the endogenous framework. Although exploratory regression models were calculated, this study can not help to distinguish between causes and consequences of e-business adoption.

Due to the highly globalised nature of the software industry, the findings should also be applicable for other countries, especially in the European Union. Future e-business research at our department will focus on in-depth case-based analysis to create an Austrian taxonomy of e-business strategies for small software business which can be compared with the wide selection of approaches, typologies as well as other taxonomies brought forward in academic literature.

REFERENCES

- [1] Z. J. Acs and D. B. Audretsch, "Innovation in large and small firms: an empirical analysis," *American Economic Review*, vol. 78, 1988.
- [2] D. Keeble, "Small Firms, Innovation and Regional Development in Britain in the 1990s," ESRC Centre for Business Research, Working Papers from ESRC Centre for Business Research 1996.
- [3] K. L. R. Pavitt, M. Robson, and J. Townsend, "The size distribution of innovating firms in the UK: 1945-83," *Journal of Industrial Economics*, vol. 35, 1987.
- [4] P. Damaskopoulos and T. Evgeniou, "Adoption of New Economy Practices by SMEs in Eastern Europe," *European Management Journal*, vol. 21, pp. 133-145, 2003.
- [5] E. W. N. Bernroider, "Micro- and small-to-medium sized enterprises in the Austrian computer software industry: Their economic importance, software process characteristics and ICT utilisation," presented at Information Resources Management Association (IRMA) 2003 Annual International Conference, Philadelphia, USA, 2003.
- [6] D. D. Phan, "E-business development for competitive advantages: a case study," *Information & Management*, vol. 40, pp. 581-590, 2003.
- [7] M. Porter, "Strategy and the Internet," *Harvard Business Review*, vol. 79, pp. 63-77, 2001.
- [8] K. Lal, "E-Business and Export Behavior: Evidence from Indian Firms," *World Development*, vol. 32, pp. 505-517, 2004.
- [9] K. Lal, "E-business and manufacturing sector: a study of small and medium-sized enterprises in India," *Research Policy*, vol. 31, pp. 1199-1211, 2002.
- [10] C. Dussart, "Transformative Power of e-Business Over Consumer Brands," *European Management Journal*, vol. 19, pp. 629-637, 2001.
- [11] J. Zheng, N. Caldwell, C. Harland, P. Powell, M. Woerndl, and S. Xu, "Small firms and e-business: cautiousness, contingency and cost-benefit," *Journal of Purchasing and Supply Management*, vol. 10, pp. 27-39, 2004.
- [12] NAICS-Association, "NAICS - North American Industry Classification System," vol. 2003, 1997.
- [13] E. S. Steinmueller, "The U.S. Software Industry: An Analysis and Interpretive History," in *The International Software Industry*, D. C. Mowery, Ed. NY: Oxford University Press, 1996, pp. 15-52.
- [14] A. C. Hax and N. S. Majluf, *The Strategy Concept and Process: A Pragmatic Approach*. London: Prentice-Hall, 1991.
- [15] C. W. L. Hill and G. R. Jones, *Strategic Management: An Integrated Approach*, 5th ed. Boston: Houghton Mifflin, 2000.
- [16] G. Houben, K. Lenie, and K. Vanhoof, "A knowledge-based SWOT-analysis system as an instrument for strategic planning in small and medium sized enterprises," *Decision Support Systems*, vol. 26, pp. 125-135, 1999.
- [17] B. N. Roth and S. A. Washburn, "Developing Strategy," *Journal of Management Consulting*, vol. 10, pp. 50-54, 1999.
- [18] K. W. Glaister and R. J. Falshaw, "Strategic Planning: Still Going Strong?," *Long Range Planning*, vol. 32, pp. 107-116, 1999.
- [19] M. Porter, "How Competitive Forces Shape Strategy," in *On Competition*, M. Porter, Ed. Boston: Harvard Business School Press, 1998, pp. 21-38.
- [20] I. Brookes and J. Weatherston, *he Business Environment - Challenges and Changes*. London: rentice Hall, 1997.
- [21] S. J. Haberman, *Analysis of qualitative data*. London: Academic Press, 1978.
- [22] S. Drew, "Strategic Uses of E-Commerce by SMEs in the East of England," *European Management Journal*, vol. 21, pp. 79-88, 2003.

Extreme Availability: Determining the Limits of Availability in Commercial Systems

David W. Coleman, Todd W. Mummert

IBM T.J. Watson Research Center, Hawthorne, NY, 10532, USA
{xinu, mummert}@us.ibm.com

ABSTRACT

High availability has become increasingly important to businesses with the advent of the Web and various associated e-commerce services. This paper describes the key attributes of the very highest availability systems. Such systems and their associated support environments are qualitatively different than other systems, even other high availability systems; for the purposes of this paper these systems at the current limit of possible availability are termed *extreme availability* systems.

Keywords: High availability, fault tolerance, commercial systems

1. INTRODUCTION

The basic requirements that any trading floor application for the New York Stock Exchange (NYSE) needs to meet are fairly easy to define. The system needs to be extremely reliable, highly available, very fast, and highly scalable. While these requirements are not unusual *per se*, the levels needed by the NYSE are well beyond what most commercial off-the-shelf software provides. The NYSE is one example of an *extreme availability* environment, where no outage during trading hours is acceptable.

Historically, extreme availability has been a specialty area, of interest mostly for space-craft systems and those systems where failure would cause significant destruction and loss of life. The usual techniques for improving availability, such as redundancy and independent backup systems, have been used with great success in such systems.

Commercial systems are extremely cost sensitive, although the degree of sensitivity varies by industry. Most businesses think about availability in much the same way that they think of disaster recovery costs – they determine the cost per hour (or sometimes per minute) of an outage, and then take appropriate steps to mitigate outages using the outage cost per hour as a cap on potential mitigation expenses. One corollary of this approach is that when business critical systems are computerized the willingness to spend to increase availability goes up sharply, since the expense cap becomes tied to the entire business revenue stream.

Given this background, it should come as no surprise that e-commerce is causing many companies to become much more concerned about high availability, and in some cases businesses are starting to look seriously at what is required to achieve extreme availability. At the moment, financial companies are most interested in extreme availability, partly because they are often the most advanced in computerizing their business

processes and partly because they tend to have the highest cost per hour for outages. It appears, however, that a secondary effect of the shift to e-commerce is that high availability is becoming a critical competitive attribute –and that extreme availability may in turn start becoming a competitive differentiator.

There are a few key ideas which appear over and over when considering the various aspects of availability. They are listed here both as an introduction to the concepts and as a convenient reference point.

MTBF stands for mean time before failure, or sometimes mean time between failures. These two phrases are equivalent only for those cases where the curve which describes the failure probability over time is a negative exponential curve, but as it happens such a curve is frequently a good fit for hardware failures, particularly for electronic or electrical equipment. In summary, MTBF is a measure of how long something is likely to stay up. Generally MTBF values are specified as some number of months or years.

MTTR stands for mean time to repair. It is a measure of how long it takes to fix something once it breaks. Note that the repair time includes whatever time it takes to diagnose the problem, acquire the repair components, pull the failing component, do whatever repair is needed, and reinstall the now working component.

SPOF stands for single point of failure. A single point of failure is a system or environmental component which, when it fails, takes down the entire system. The significance of a SPOF is that the reliability of that component will be an upper bound on the overall reliability of the system. In general, any SPOF is a weak link in a high availability chain.

This paper provides a background in the general approaches and concerns for very high availability systems. While these general approaches are well known in certain areas, it is useful to revisit them in the

context of distributed e-commerce systems to understand what aspects are most important in that environment. To do this, the paper follows the general overview with an overview of the techniques and approaches used by those organizations which are most aggressively pursuing extreme availability. Finally, some general information is given from case studies of such extreme availability systems, to indicate what the limits of the commercial state of the art are today.

2. FAULT TOLERANCE

Fault tolerance is the study of how to design systems or subsystems so that they continue to operate correctly in the presence of components which fail. Normally the measure for a successful design is an increase in the overall system MTBF. There are some curious tidbits from this area of study, such as the fact that redundancy can actually *decrease* system MTBF. To understand how and why such things can happen, it is necessary to understand how component MTBF can impact system MTBF. [1] [2]

2.1 MTBF –Underlying Assumptions

Large systems are composed of many parts, and to compute the MTBF for such large systems we need a way of composing the MTBFs for the parts. In general,

$$\begin{aligned} &P(\text{exactly one of } E_1 \text{ and } E_2 \text{ occurs}) \\ &= P(E_1) + P(E_2) - (P(E_1) * P(E_2)) \end{aligned} \quad (1)$$

Note that for sufficiently small probabilities, the product term will be quite small with respect to the overall value, and hence we can approximate the probability that exactly one of E_1 and E_2 occurs with the sum of the probabilities of E_1 and E_2 . (Since for fault tolerance we are dealing with failure probabilities, and most failure probabilities for IT systems are less than 5%, this simplification is useful and fairly accurate. Note that $5\% * 5\%$ is 0.25%, while the sum is 10%, and that further the accuracy of the approximation improves as the probabilities get smaller.)

Because we are interested in the expected life span (Mean Time Before Failure), rather than the probability of failure in a given time span, we must find a new function of $P(E)$ which yields this information. For random probability distributions this can be quite difficult, but it turns out that a useful approximation is to assume that the probability is memory-less, which means that the probability of failing in the next time interval is independent of how long it has already been in service. With that assumption, we get

$$MTBF(E) = 1/P(E) \quad (2)$$

It turns out that many IT devices actually exhibit a bathtub curve probability, with a high infant mortality rate and a high wear out rate joined by a long period of essentially constant and low mortality. Manufacturers

can convert such devices into memory-less devices by using a burn-in period to eliminate infant mortality, and setting the nominal end of life (or end of warranty) prior to the rise due to end of life mortality.

For a device with a 5% probability of failure in any given month, we will therefore get a 20 month MTBF. Most systems are composed of many modules. If the modules fail independently, we can compute the probability of the system failing. In general, for any set of N identical parts the MTBF for the entire set is $1/N$ times the MTBF for one such part.

2.2 Failure Modes

Since we are interested in fault tolerance, it is natural to ask about the failure modes of the various devices we are considering. Unfortunately, this question is extremely hard to answer with complete accuracy for most devices. Further, the answer to such a question is tightly tied to the exact technologies used, so even having the answer for a given device doesn't necessarily help in determining the answer for another similar but not identical device.

There are a few general principles which can help increase fault tolerance. The first such principle deals with what are commonly called soft, or transitory errors. Soft errors have the property that the probability of occurrence is independent of the work being done, and further they do not always occur even when the same work is done. There can be various causes for soft errors, but in terms of fault tolerance the ultimate goal with soft errors is always the same – to mask the occurrence of the error in such a way that there is no immediate fault generated. The solution in these cases always comes down to redundancy, and typically it is implemented via error correcting codes. While ECC is a large and complex field of study, it is also sufficiently well understood that such codes are in widespread use by device manufacturers to control transitory errors. Such codes are used in all disk drives, and on almost all communication channels.

Workload independent hard errors (also known as non-transitory or permanent) can also be handled by manufacturers by using redundancy, but in this case the redundancy needed is additional hardware.

Unfortunately there are several additional classes of errors which do not easily lend themselves to general solutions. Workload dependent errors, such as the floating point problem in Intel processors, are one major example. Additionally, many devices have partial failure modes, where things are neither working completely nor failed completely. While most manufacturers attempt to identify such problems with testing, even the most rigorous testing will occasionally miss such problems.

There is one final general approach to dealing with failure which can be helpful, both in the real world and for theoretical analysis of fault tolerance. This approach is called failstop, and it works by forcing all faults into the same terminal state, namely the stopped state. This approach significantly aids in the design and operation of fault tolerant systems, and as will be seen below it also aids in the analysis of the behavior of such systems.

2.3 Fault Tolerance Approaches

In general, fault tolerance requires two things, namely a way of detecting failure and enough redundancy to allow the failure to be overcome.

Under certain conditions, it is possible to turn any device into a failstop device. The simplest way to do that with non-failstop components is to duplicate the device, and then use some kind of comparator to see if the outputs of the two copies of the device agree. If the outputs agree, the device continues to work. If the outputs disagree, the device stops. For this process to be feasible, of course, it must be possible to build the needed output comparator. The major drawback to this simple form of failstop creation is the impact on MTBF. Because there are two copies of the basic device, the failstop device has an MTBF which is no more than $\frac{1}{2}$ the MTBF of the original device. This can of course be mitigated by using more copies of the device; with three copies (and a more sophisticated comparator) the failstop device has an MTBF of $\frac{5}{6}$ of the MTBF of the original device. (An additional advantage of triplex or higher redundancy arises if soft errors predominate in the failure modes for the base components. In that case the comparison process will completely eliminate the soft errors, and significantly boost the overall MTBF.)

Although building failstop devices doesn't sound obviously attractive at this point, things become better once higher layers are built on top of such failstop devices. With non-failstop components, the device must necessarily fail as soon as there is no longer a majority of devices working. With failstop components, the device can continue to work as long as at least one of the components is working, since each component will either work correctly or stop. While this won't ease the redundancy cost of the additional hardware, it has a substantial impact on MTBF. With a duplex device, the overall MTBF moves from $\frac{1}{2}$ the component MTBF to 1.5 times the component MTBF. With a triplex device, the overall MTBF moves from $\frac{5}{6}$ the component MTBF to $\frac{11}{6}$ times the component MTBF.

The most interesting advantage of building with failstop components only appears with one additional assumption. If it is possible to replace failed components without taking down the entire device, and

synchronize the new component if necessary, then the MTBF calculations change significantly. The device will fail if all but one of the components are unavailable, and then the remaining component fails. This yields the following equation for a device made with n copies of a failstop component:

The MTBF for the n component device with repair is:

$$MTBF/n * (MTBF/MTTR)^{n-1} \quad (3)$$

or rewriting slightly

$$MTBF^n / (n * (MTTR)^{n-1}) \quad (4)$$

Using our usual 95% reliable components, and assuming that repair takes a full day, we find that the MTBF for a duplex device is made from 20 month MTBF components is approximately 1000 months, or roughly 80 years. Using the same components, a triplex device will provide an MTBF of approximately 80,000 months, or roughly 6500 years.

Note that using a comparator and duplex or triplex (or more) is not the only way to provide fault tolerance. The key to fault tolerance is spare hardware and some way to detect failure, with the spare hardware used to correct the detected fault. As an example of another alternative approach, consider RAID 5 disk drives. What is done for RAID 5 is to use a redundant coding on the, with the extra drives needed by RAID 5 used to increase storage to hold the extra bits required by the redundant encoding. [3] [4] [5]

2.4 The Bad News

From the preceding discussion, it is clear that whenever we have workload independent failure modes, either soft or hard, and devices which are memory-less in their failure mode, which are relatively reliable and easy to repair, it is possible to build fault tolerant devices with very high MTBF.

An obvious question is then what large system components do not exhibit the needed properties. And unfortunately there is one key area which does not fit the picture—software. Even worse, the characteristic which is least compatible with the assumptions is the primary assumption behind much of the fault tolerance work, namely the assumption that most failure modes are workload independent.

When most of the failures are a function of the workload done, no voting scheme which uses multiple copies of the same software increases reliability.

As a practical matter, most people designing systems for high availability deal with the issue of software faults by relying on one of the demonstrated advantages of software, namely that the defect rate decreases with use. Thus, it is common for systems with high sensitivity to failure to prefer software which has been in wide or long term use elsewhere to take advantage of the

improved defect rate. Additionally, there is likely to be a much greater emphasis on testing for such systems.

One additional area of workload dependent failure warrants special attention currently. This is the area of security issues and faults. While there have been security issues and bugs since the earliest days of computing, even in systems which were designed from the ground up with security in mind, the Internet has increased the significance of such issues. The problem is that the Internet makes networked computers more widely available. While this is a good thing in terms of increasing the usefulness of such computer systems, it also allows those with harmful intentions access to the systems. These people will typically search out security faults of various types, in order to exploit the weaknesses for their own purposes.

2.5 Fault Tolerance Summary

Standard failure analysis has yielded tremendous gains in hardware component reliability (disk drives have gone from MTBF of hundreds of hours to MTBF of tens of thousands of hours over the last few decades for instance). Further, the work on high availability systems has yielded a number of highly effective techniques for composing the improved components into larger systems which are even more reliable (one notable innovation being RAID 5).

The approaches used for hardware fault tolerance have not been noticeably successful in providing software fault tolerance. Curiously enough, until recently this has not been a major issue in the design of highly available systems, largely because the software failure rate was sufficiently low that other modes of failure dominated in the overall failure modes. Improvements in hardware reliability, coupled with a trend toward attacks on systems via software defects, are likely to increase the urgency for solutions in this area going forward.

3. HIGH AVAILABILITY

High availability is yet another way of looking at the issue of IT system faults and the key ideas of MTBF and MTTR. In the case of high availability, the focus is on the end user experience. Thus, high availability is normally expressed in terms of per cent availability of the process or function supplied by the associated IT system. Note that availability will be increased by either increasing MTBF or by reducing MTTR

3.1 Availability and Target Percentages

While availability is normally expressed in percentage terms, since those percentages are frequently over 90%, it can be useful to convert the percentages into minutes of downtime per year. Thus 99.9% availability is about an hour a year of downtime, while 99.999%

availability is about 5 minutes a year of downtime.

While this is easy enough to understand, there is a complication in how availability can be calculated which can impact the downtime numbers above. In computing downtime, some people exclude planned downtime (for things like upgrades or hardware maintenance), and some don't.

The mathematical calculations above are interesting. More interesting is the question of what end users should really expect. In general this is a hard question, which depends on many details of the actual system under consideration, but the following quote seems pertinent.

The availability specification for the Windows NT server is about 99.5 percent. UNIX has something like 99.7 percent availability and OS/390 has 99.999. When you think about it, that means your unplanned downtime on an NT server will average about 50 hours per year. With UNIX, the number is around 26 hours per year. And, with OS/390, you are looking at 5 minutes of unplanned downtime per year. Considering the cost to companies for an hour of unplanned downtime -- \$100,000 for a retail operation, \$6 million or more for a brokerage -- you can see that the drive to client/server is a costly one. [6]

3.2 Improving Availability

Improving availability is not an all or nothing process. It is possible to improve the availability of individual subsystems. This is fortunate, since the cost of applying high availability principles to entire IT systems will increase the cost of the system by a factor of 2 to 4, depending on system details.

The obvious place to start in improving availability is wherever outages are causing the most pain. For each such pain point, alternatives can be developed using the various techniques which have been outlined above. Then the cost of implementing the various alternatives can be determined, and weighed against the business benefit to be gained. As an example, let us assume disk drive failures are causing a significant productivity loss. A few alternatives to increase availability are backups (which can significantly reduce MTTR), raid arrays, various kinds of networked storage (SAN or NAS), and of course using a more reliable brand of disk drive. The disk drive business has grown sufficiently competitive that all the major vendors offer roughly equivalent MTBF ratings, so that alternative is unlikely to bring useful gains in reliability. Backups are likely the cheapest alternative, so if an hours long outage for restoration (and possible loss of any work done since the last backup) is acceptable, this might be a worthwhile alternative. Raid arrays will entail the cost of a new raid controller and additional disk drives for the particular raid mode chosen, but will provide

significantly increased MTBF, and in some cases lower MTTR also. Networked storage will entail acquiring the appropriate storage boxes, as well as some increased administrative overhead. Note further that both networked storage solutions and backup solutions can be shared among multiple systems, so even determining the cost of an alternative is specific to the business making the choice. Given the multitude of technically feasible solutions, business requirements and tradeoffs dominate in determining the exact nature and level of high availability appropriate for any given IT system.

Any business with a large number of similar systems can usually determine fairly easily which components most commonly cause outages. For smaller businesses, or for unusual configurations, it will be necessary to rely on outside data to determine the components most likely to cause problems.

One final point may be useful for any business working on improving availability. Initially it is easy to look at overall statistics on business wide outages and determine which areas need attention. However, once availability improves enough, even a large business is likely to find that the weekly variation in outage causes starts to disguise the underlying trends, and make it hard to gain further improvement. The exact point at which this occurs is largely a function of the size of the business, but in general this sort of problem will start to appear around 99.5% availability. When this starts to happen, it will likely be useful to shift to considering not overall outage statistics, but rather statistics on the longest outages. Normally at this point reducing the length of the longest outages will provide a significant boost to overall availability. [7]

3.3 An Overview of System Failures

As we have seen above, having information on what kinds of system or component failures to expect makes the process of improving availability much easier. This makes published data summarizing all failures, by source of failure, for a large number of systems of considerable value. There are, however, two problems with such data.

The first problem is that there are almost no such studies. The only published work we are aware of that cover all system failures for a large number of systems is the work covering an analysis done by the Japanese Information Processing Development Corporation (JIPDEC) in the mid 1980's. This study surveyed the outages reported by 1,383 businesses in 1985, and covers 7,517 outages with an average outage duration of about 90 minutes. This yields an overall MTBF for the systems of about ten weeks, and an availability of 99.91%. [8]

The second problem is that the data is very dependent on the technologies used in creating the system. While

there are specialists who can determine the expected reliability of a system, using detailed calculations on all the various components of the system, this work is quite hard and expensive – normally it is easier to simply observe actual performance. This is particularly true since one of the key components of the system, the software (operating system, middleware, and application) typically does not come with a rigorous MTBF guarantee.

Despite these two caveats, it is interesting to see what has been found, and combine it with the technology trends in the intervening years to provide some guidance on what areas will most likely benefit from improved availability efforts.

In the mid-1980's conventional wisdom was that disks were the major concern for reliability, because they were mechanical. This was borne out by disk MTBF values of about 10,000 hours, although some vendors were producing disks which were both significantly more expensive and more reliable. Although disks have remained mechanical, they now exhibit advertised MTBF values of 50,000 hours, and actual experience indicates MTBF values of between 5 and 10 years. While not all the MTBF news is good, as will be seen in the next section, in general computer hardware has become significantly more reliable, and requires significantly less maintenance than was true during the JIPDEC study.

But is improved hardware and maintenance enough? Note that while the MTBF with 1980's hardware and software was 10 weeks, even with perfect vendor hardware and software the JIPDEC average MTBF would only rise to 4 months. With an average repair time of 90 minutes, this would correspond to an availability of only about 99.95%. Vendor product improvements are clearly important, but also clearly not enough to achieve the high availability goals.

3.4 External Factors

It is useful to consider the impact of WAN connectivity and the power network, which are typically not under the direct control of the business attempting to provide a high availability system. In North America, communication line suppliers will provide service guarantees of 95% error free seconds, bit error rates of no more than 1 error per million bits, and 99.7% availability. Redundant message encoding and duplexing transmissions via independent paths can mask most transient failures, but 99.7% availability is significantly lower than even a PC server. Further, a 1998 study by RTO West, a Canadian power supplier indicated that for their system power was available at the point of delivery only about 98% of the time. Once again, this sort of low external availability can be overcome with careful system design.

3.5 Fault Tolerance Summary

There are five normal sources of outage: 1) Hardware; 2) Software; 3) Operations; 4) Maintenance; and 5) Environment. (Environment includes all items which are not considered part of the system of interest, such as the building(s) the system sits in or the power or machine room cooling systems.)

The base reliability of computer and network hardware has been improving for some time. One recent study suggests that even with PCs, which are normally considered the least reliable server alternative, it is possible to achieve 99.5% availability and an MTBF of about 35,000 hours. While this may change with increasing heat stress in newer PCs, it is still a very promising starting point.

On the other hand, there is no evidence that software reliability is improving anywhere near as fast. Further, virus attacks, which rely on defective software for their effects, seem to be increasing at a rather startling rate. This suggests that for the foreseeable future the limiting factor for availability is likely to be the software.

Operations and Maintenance outages are typically the result of human error and/or vendor design problems in tools or procedures.

Environmental outages include all the sorts of issues that are considered as part of a disaster recovery plan. It is clear from looking at just a few of these items that if they are not taken into account then environmental issues will provide an upper bound on availability significantly below 99%. On the other hand, with careful planning most or all such outages can be controlled.

It is worth noting that high availability is not an all or nothing attribute. By addressing any of the normal sources of system failure, the overall availability will improve. Similarly, by improving the MTTR for any source of failure, availability will improve. For most businesses, the key to improving availability is gathering the data which will cost justify the needed system changes.

4. EXTREME AVAILABILITY

Extreme availability is the ultimate form of high availability, when the desired goal is no outages at all, because the perceived harm to the business is immeasurably large. How does this differ from high availability? To answer that, a comparison is useful. In the United States, both the nuclear power industry and nuclear powered naval vessels were expertly designed with careful attention paid to fault tolerance. Further, in both instances the systems are run by thoroughly trained experts. Nonetheless, the nuclear power industry has had several "incidents", while no

such incidents have been encountered in the nuclear navy.

4.1 Critical Components

In even the most carefully designed fault tolerant system, there will be unexpected events. While such events will be rare with careful planning, proper handling is vital if outages are completely unacceptable.

In addition to the importance of people, most extreme availability organizations pay close attention to a number of other areas in order to achieve their results. These areas include culture, fault tolerance, upgrade procedures, testing, security, using mature software, documentation, and simplicity – each of which is covered in more detail in the following paragraphs.

The first common feature of extreme availability businesses is that they have a strong culture of reliability, which emphasizes such things as always taking extra care and always being ready for a surprise. While culture is always somewhat specific to the organization if appears in, extreme availability cultures seem to share certain traits.

As we have seen in earlier sections, there are a great many possible sources of problems. Extreme availability requires a fanatical attention to possible single points of failure. In power, for example, the building should have two power feeds, coming in on opposite sides of the building, from separate power substations. Each power feed should feed an independent electrical distribution network inside the building. The outside power should be backed up by batteries, which in turn are backed up by generators, with an on site fuel storage sufficient to last many days. Good planning might even get multiple fuel delivery contracts (with independent providers) which would start a week or so before the fuel store would be emptied.

One area where extreme availability organizations pay particular attention to fault tolerance is in the upgrade process for hardware and/or software. For upgrades, extreme availability organizations use a gradual rollout process of some sort, which allows the organization to observe the upgrade in actual operation. Further, this gradual rollout is coupled with a way to back out the upgrade if problems occur. Note that being able to do gradual rollout and immediate back-out (if necessary) may well impose additional requirements on the design for system redundancy.

Redundancy and fault tolerance can protect against random errors, but they provide no protection against systemic errors, such as a programming mistake or misconfigured hardware. The only remedy for such systemic problems is testing. Extreme availability organizations take testing very seriously. Testing

periods are never shortened. Frequently there are multiple independent testing organizations, and each of them has the ability to stop deployment. Testing is normally not under development.

In addition to systemic threats, there are the active threats from people who for whatever reason want to impact the system. Security precautions will include both physical security (site, building and machine room) and network security. When possible, the system should be isolated from the Internet. If Internet access is a key feature of the system, then layered protection against penetration attacks are needed, as well as a mechanism for handling denial of service attacks.

Security precautions are important, of course, but beyond that extreme availability organizations typically implement a higher than normal degree of isolation. For example, development networks may be fully independent of production networks. And of course testing environments are fully independent of both development and production.

Software is markedly different from hardware in the shape of its failure curve over time. New software is markedly more likely to fail than software which has been used by many organizations over several years. Many organizations recognize this effect by avoiding all x.0 releases. Extreme availability organizations extend that rule to try and avoid any commercial software until it has a track record with other organizations.

As we have seen from earlier sections, high availability relies on fault tolerance plus repair. Good monitoring is key to understanding when there is an otherwise transparent failure, and hence when repair is needed. A correct view of what is happening in the system also lowers the likelihood of an erroneous action being taken by an operator.

Extreme availability requires an unusually complete set of operational documentation. Why? People can be single points of failure also. Worse, even if you have several people cross trained you may find that the backup is out sick while the primary is on vacation.

Simplicity is a watchword for most extreme availability organizations. Simplicity means using the fewest pieces possible for the function needed, because of the impact on MTBF of additional components. It also means using simple administrative interfaces, to lessen the chance of error in an emergency.

4.2 Extreme Availability Summary

Extreme availability takes a high availability system as a base, and adds operations personnel with good skills handling emergencies and a strong culture of failure avoidance.

While this may sound simple, it is not. Any culture change is difficult, and the culture of high availability is not common in the industry. This means that it is a difficult adjustment for many.

5. CONTINUOUS AVAILABILITY

As the Internet becomes more significant to business, increasing numbers of firms are looking at the issue of making some or all their systems always available. eBay, for example, attempts to make their systems always available. For convenience, this type of availability is called continuous availability. Given the success of the various approaches above in increasing availability, an obvious question is what impact the additional constraint of continuous availability might have.

At first blush, it would seem that there should be very little impact produced by the continuous constraint. After all, the most highly available systems are already up almost all the time. But in fact there is a substantial impact, and the reason is contained in one of the points touched on earlier, namely the variations in the definition of availability. Most very high availability systems have availability defined in terms of scheduled uptime. Or to put it another way, they do not include scheduled downtime in their availability calculations.

5.1 Implications of No Scheduled Downtime

As we saw in the section on fault tolerance, to get the highest levels of availability, it must be possible to repair a system while the system continues to operate. Given this, it seems it should be possible to achieve the same levels of availability for continuously available systems as for those that have some small degree of scheduled downtime.

The reason this is not in fact the case is that manual repair time is actually quite dangerous from a system availability point of view. The data in this area is quite hard to come by, and is also very fragmentary, but it appears that outages are approximately 30 times more likely during a repair session. This is of course a function of human fallibility. What appears to be happening is that, since repairs are a rare event, people are much more likely to make mistakes at that time than at other times. With scheduled downtime, such mistakes can be caught and fixed without impacting availability.

There is of course a considerable body of data on human errors during emergencies. (A continuous availability outage is naturally an emergency.) It is interesting to speculate on whether well designed repair procedures, which are protected against human error, could in fact allow continuous availability to be a trivial byproduct of high availability. Current hardware and software upgrade and repair procedures do not allow this. As a

result, companies such as eBay, which aim for continuous availability, have an availability target of 99.9%, while firms which aim for extreme availability are likely to have a target of 99.999%.

5.2 Continuous Versus Extreme Availability

Despite the fact that continuous extreme availability appears to be beyond the current state of the art, continuous availability and extreme availability still share a number of similarities.

Both continuous availability and extreme availability take as their base careful attention to high availability principles. Further, both rely on skilled operations personnel working in a culture which emphasizes availability.

The same cultural focus items which are important to extreme availability tend to be important for continuous availability, although there are a few differences which will be detailed below.

While continuously available environments need the same degree of fault tolerance as extreme availability environments, there is an additional requirement. The additional requirement is that repairs must be possible to a running system. This is a design constraint on the system, and it applies to all parts of the system. Notice for example that this can imply a need to be able to replace a power outlet without impacting any other outlet on the same circuit.

Gradual rollouts are important for both extreme availability and continuous availability, but for continuous availability the rollout process (or rollback process) must occur while the system is live. Once again this is a design constraint, but in this case the design constraint may impact such things as the software communication protocols, since a gradual rollout will create situations that could otherwise be avoided.

Security needs are similar in extreme availability and continuous availability, but for continuous availability it will be necessary to have security upgrades rolled out on a live system.

Monitoring is somewhat more complex for continuously available systems, since monitoring will be needed for the repair process also.

5.3 Continuous Availability Summary

At the moment, the very highest levels of availability appear to be unachievable in a continuous availability setting. Careful attention to the high availability precepts, combined with an attentive workforce, can achieve about 99.9% availability, or roughly 9 hours of outage per year.

6. CONCLUSION

While great strides have been made in improving IT hardware availability and in understanding the design principles and operating procedures which will aid in increasing availability, it remains true that high availability is neither easy nor free.

Well designed redundant hardware operated by skilled practitioners can achieve availability rates of 99.999% for key periods each day, or 99.9% on a continuous basis. Organizations which achieve such values spend significantly more on their IT infrastructure than the average commercial enterprise. Unfortunately as we move towards greater degrees of electronic integration between businesses, there will be a need to both close the gap in continuous availability and a need to make extreme availability feasible for businesses that do not have a large team of highly skilled operators.

REFERENCES

- [1] Avizienis, A., H. Kopetz, J. Laprie. *Dependable Computing and Fault-Tolerant Systems*. New York: Springer-Verlag, 1987.
- [2] Bowen, J. P., M. Hinchey. *High-Integrity System Specification and Design*. London: Springer-Verlag, 1999.
- [3] Patterson, D., J. Hennessy. *Computer Architecture: A Quantitative Approach*. San Mateo, California: Morgan Kaufmann, 1990.
- [4] Pradhan, D. *Fault Tolerant Computing: Theory and Techniques*. 2 volumes. Englewood Cliffs, New Jersey: Prentice Hall, 1986.
- [5] Siewiorek, D., R. Swarz. *Reliable Computer Systems, Design and Evaluation*. 2nd edition. Bedford, Massachusetts: Digital Press, 1992.
- [6] Toigo, J. W. *Disaster Recovery Planning: Preparing for the Unthinkable*. 3rd edition. Upper Saddle River, New Jersey: Prentice Hall, 2003.
- [7] Marcus, E.; H. Stern. *Blueprints for High Availability*. New York: John Wiley and Sons, 2000.
- [8] Gray, J., A. Reuter. *Transaction Processing: Concepts and Techniques*. San Francisco, California: Morgan Kaufmann Publishers, 1993.

Impact of Internet Banking on Customer Satisfaction and Loyalty: A Conceptual Model

Thanh Nguyen, Mohini Singh

School of Business Information Technology, RMIT University,
GPO Box 4576V, V Melbourne, Victoria, Australia
thanh.nguyen@rmit.edu.au, mohini.singh@rmit.edu.au

ABSTRACT

This is a research in progress paper discussing Internet banking issues both from the bank's point of view and from the customers' points of view. Based on literature it includes several factors that impact Internet banking. It includes several hypothesis derived from literature to be tested empirically. The research will be guided by the conceptual model presented as figure 2 in this paper highlighting the relationship between Internet banking and customer satisfaction and loyalty.

Keywords: Internet banking, customer satisfaction, customer loyalty

1. INTRODUCTION

Internet banking is one form of e-commerce that has gained a wider acceptance than others. In Australia all banks are now offering internet banking facilities to all its customers. As suggested by Chang [10] the evolution of information technology has significantly influenced the banking industry. Particularly, the advent of the Internet and the popularity of personal computers have created both an opportunity and a challenge for this industry. In only a decade Internet Banking has rapidly grown [2]. In 2001, the Market Intelligence Strategy Centre (MISC) reported that approximately 2.8 million people have used Internet Banking in Australia [50]. These figures increase rapidly with each new survey. A study from Taylor Nelson Sofre in 2002 revealed that 23% of Australians used Internet Banking and the Internet Banking adoption rate in Australia is greater than in the US and Britain [3]. In mid-2002, the MISC reported that 5 million Australians had used Internet Banking services [39]. Recent survey undertaken by the MISC shows that the number of Australians using Internet Bank had reached 7.7 million in September 2003, with an annual growth rate of 113% [33].

The rapid growth of Internet banking in recent years is a clear indication of consumer acceptance of this media. However, the impact of Internet Banking on customer satisfaction and loyalty in Australia is yet to be established as few studies have addressed this issue. In this paper we present a review of literature on internet banking and the factors that may impact internet banking and customer satisfaction and loyalty.

2. DEFINITION OF INTERNET BANKING

Internet Banking is defined as the usage of Internet and telecommunication networks to deliver banking services to customers [12, 14, 43]. Customers can inquire information and carry out most banking services such as account balance inquiry, inter-account transfers, and bill-payment via the Internet. There are different

perceptions of Internet Banking between the literature and practitioners. With reference to most Australian banks' websites, the term 'Internet banking' has been construed as the transactions relating to current and credit card accounts such as viewing balances, paying bills, and transferring funds. In contrast, in the literature, Internet banking includes the services relating to financing, insurance, investment, and new banking services [14, 43]. The latter will be referenced in this paper.

3. INTERNET BANKING TECHNOLOGY

According to Leong, Srikanthan & Hura [27] Internet banking is a web application based on client/server architecture in which Internet technology plays an important role. A basic architecture is illustrated in figure 1.

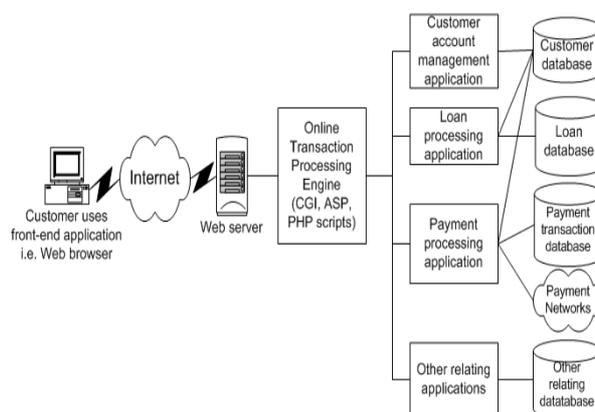


Figure 1: Internet Banking application architecture - adopted from Leong, Srikanthan & Hura (1998) and Turban et al. (2000)

Similar to all Internet transmission of information, with Internet banking security is an important concern for banks and customers [43]. All banks offering Internet Banking have taken special care to ensure security, privacy and confidentiality of information to all its customers. Basic security requirements to conduct business over the Internet are discussed below. *Authentication* involves the ability of an individual, organisation, or computer to prove its identity. Security systems accomplish authentication by verifying information that the user provides against what the system already knows about the user. *Authorisation* involves the control of access to particular information once identity has been verified. Authorisation is meant to limit the actions or operations that authenticated parties are able to perform in a networked environment. *Audits* include information on access of particular resources using particular privileges or performing certain security actions. It identifies the person or program that performed the actions. *Confidentiality* involves the secrecy of data and/or information, and the protection of such information from unauthorised access. For e-businesses confidentiality is of utmost importance in the protection of an organisation or company's financial data, product development information, organisation structures and various other types of information. Time related information such as a price list or confidential report can be crucial and need to be kept confidential until a certain time. Policies regarding the release of information must be included in confidentiality, as well as authorisation services. Confidentiality policy must ensure that information cannot be read, copied, modified, or disclosed without proper authorisation and that communication over networks cannot be intercepted. *Integrity* is the protection of data from modification either while in transit or in storage. E-commerce and e-business systems must have the capability of ensuring that data transmissions over networks arrive at their destinations in exactly the same form as they were sent. Integrity services must protect data against modifications, additions, deletions, and reordering parts of the data. Information can be erased or become inaccessible, resulting in loss of *availability*. This means that those who are authorized to get information did not get what they needed. Availability of information is an important attribute in service-oriented businesses that depend on information (e.g. airline schedules). When a user cannot get access to the network or specific services provided on the network, they experience a *denial of service*. *Nonrepudiation* involves protection against a party involved in a transaction or communication activity that later falsely denies that the transaction or activity occurred. Nonrepudiation services must be able to demonstrate to a third party proof of origin, delivery, submission and transport of the data in question [56].

Security protection includes physical security devices i.e. secure server rooms, unauthorised intrusion detector, organisation's security policies and related Internet security technologies. Authentication, digital

certificates, firewalls, Secure Socket Layer (SSL), Secure HTTP (S-HTTP), Secure Electronic Transaction (SET), Joint Electronic Payment Initiative (JEPI), and Open Financial Exchange (OFX) are important technologies supporting Internet Banking security [25]. SSL and S-HTTP are popular protocols for secure communications over the Internet at network level security and application level security respectively, using public key cryptography and symmetric key encryptions. On the other hand, SET is a standard/protocol for securing real-time credit card transactions over the Internet. This protocol enables E-Commerce applications to communicate between the customers, SET-compliant merchants and authorised third-party payment gateways.

JEPI and OFX are secure communication protocols. JEPI is known to standardise payment negotiations by assisting consumers and merchants in selecting an appropriate payment mechanism of different payment protocols for both parties. OFX is primarily designed for electronic data exchange between financial institutions and is a messaging protocol for online financial services such as Electronic Bill Presentments and Payments (EBPP) and investment services. OFX has been adopted by many large financial institutions [25]. In Australia, OFX has been implemented for BPAY, an EBPP system developed by Australian banks [9].

Internet Banking allows account aggregation. With account aggregation, customers can manage several banking accounts held at different banks via a unique access point. This means that customers just log on to the Internet Banking websites which support account aggregation enable customers to access their account held at other banks, including deposit account, credit cards and award programs [28]. To use account aggregation service, a customer has to give settings such as the URL, user name and password of each external account. Once customer logs on to an account aggregation website, this system automatically logs onto other banks' Internet Banking sites just as if a customer is logged on to them directly. With account aggregation, a customer is able to access all information from an account held at another bank. This provides customers with convenience and time saving. Although this has many advantages to the customers, security, level of information access and legal standards to protect the banks and customers has led to a slow adoption [1, 28].

In Australia, account aggregation was offered by WestPac, Commonwealth Bank, AMP and Macquarie Bank [57]. However, recently, Macquarie Bank has removed this feature due to the security concerns [47].

3. BENEFITS OF INTERNET BANKING

Internet Banking provides clear advantages to both the financial institutions and the customers. From the banks' perspective, Internet Banking has very low cost transactions, compared to human teller banking since it

reduces the following expenses [24, 52]: (1) Banks can reduce customer service staff as customers use more self-service functions; (2) There is less cheque processing costs due to an increase in electronic payments.; (3) Costs of paper and mail distribution are reduced as bank statements and disclosures are presented online; (4) There is less data entry as applications are completed and processed online by customers. On the other hand, according to KPMG [24], bank's revenue increases from Internet Banking due to: (1) Increased account sales; (2) Wider market reach; (3) New fee-based income; (4) New market opportunities; (5) Improved customer satisfaction. For consumers, Internet banking provides convenience, lower service charges, more accessible information about bank accounts, and an attractive option for busy people since it saves time to go to the bank branches and gives 24 hours access [26]. All the benefits of B2C e-commerce such as 24*7 bank service, convenience, access from anywhere, one stop shop and easy access to information [57, 58 59] also apply to internet banking.

6. PRODUCTS AND SERVICES OFFERED ONLINE

Internet Banking provides customers with many services and several facilities to carry out online banking transactions. The basic services include opening a bank account, printing bank statement, transferring funds between bank accounts, and paying bills. Other Internet Banking services comprise of future, dated and recurring bill payments and funds transfers, loan application, e-trade, fund management, customisation of account names, ability to export banking transactions to popular financial management software packages such as Microsoft Money or Intuit Quicken [1]. In addition, account aggregation services allow customers to manage multiple accounts from different financial institutions with a single login. There are many online financial services but they can be divided into the following categories [18]: (1) Current account; (2) Insurance-based; (3) Credit-based; (4) Investment-based services.

Most Australian banks provide interactive, extensive and sophisticated Internet Banking services [6]. However, it appears that no empirical study has been conducted to examine whether the Internet banking services in Australia are satisfactory to customers.

7. INTERNET BANKING AND CUSTOMER BEHAVIOUR IN THE LITERATURE

In the literature, scholars have focused on examining the characteristics of Internet Banking and the factors

In figure 1, three groups of variables that are likely to have an impact on customer satisfaction are grouped together. These are Internet Banking system quality, customer characteristics, and customer's system constraints. However, for the purpose of assisting practitioners in building a more effective Internet

influencing Internet Banking customer behaviour/attitude. The recent studies have examined the determinants such as demographics [7, 23, 32, 49, 15, 23, 32, 49, consumer characteristics [15, 23, 32, 35, 44, 34, 37, 43, 30, 45, 51], consumer perceptions [15, 44], technology attributes [30, 45, 51], and social context [15, 44] in association with the affects of these variables to Internet Banking customer behaviour/attitude.

Unfortunately, most studies concentrated on investigating the impact of the above determinants on Internet Banking adoption. Few studies [34, 40, 48, 54] have examined customer satisfaction and loyalty in relation to Internet Banking. Though Internet Banking in Australia began in 1997 [43], literature and research have only concentrated on Australian Internet Banking adoption issues, focusing on either retail customer [43] or small and medium business [52]. No empirical study has been conducted to understand the impact of Internet Banking on customer satisfaction and loyalty in Australia.

8. A CONCEPTUAL MODEL AND HYPOTHESES

For the purpose of understanding the factors influencing Internet Banking customer satisfaction and loyalty, this paper proposes a conceptual model (figure 1). This conceptual model is developed based on several previous studies relating to Internet Banking and human-computer interactions [55].

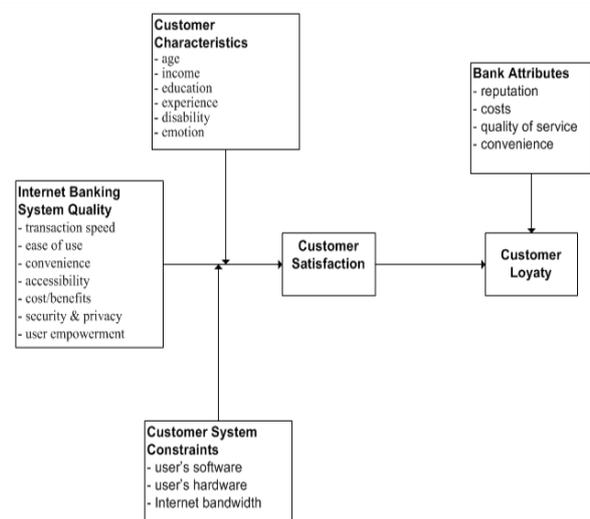


Figure 2: Conceptual model of the relationships between Internet banking system and customer satisfaction and loyalty

banking application and the impact of the technological characteristics of Internet Banking are specially considered.

Internet Banking system quality: this consists of several independent variables such as transaction speed, ease of use, convenience, accessibility, cost/benefit,

user empowerment, security and privacy [30, 34, 43, 46, 54].

Customer characteristics: Characteristics of Internet banking customers identified from literature are younger, high income earners [17, 41, 46, 17, 41], and the better-educated [17, 41]. To date no significant relationship between e-Banking and gender or occupation has been identified [17, 41]. Zhang & Li [51] suggest that health condition (disability, emotion i.e. stress) also impact customer perceived usage of technologies. Based on the above, age, income, education, experience, disability and emotion are included as customer characteristics in the conceptual model.

Customers' system constraints: System constraints (hardware, software and Internet bandwidth) are likely to affect the accessibility of Internet banking by customers [51]. Customers with these constraints will be deprived of Internet banking opportunities therefore these technology issues are considered as contingent variables.

Customer satisfaction is defined based on the *confirmation/disconfirmation* theory that states "satisfaction results from a process of comparison where consumers judge product satisfaction against their expectations about product performance"[17].

Customer loyalty, in the behavioural approach, has often been construed as repeat purchase frequency [22], probability of future purchase [36] and intentions of switching brand [19]. Cunningham [11] assesses customer loyalty based on the proportion of money that a customer spends with purchases from a single supplier. In the attitudinal approach, a loyal customer must have strong "attitudinal commitment" to a brand [16, 20].

Bank attributes: Customer loyalty/retention may result from a bank's reputation (brand name, well-known, popularity), costs (interest rate, fees), convenience (i.e. location and distance) and quality of banking service [13]. Therefore, these factors have been taken into account to identify the relationship between customer satisfaction and customer loyalty.

From the literature review and the conceptual model, the following hypotheses are thus stated:

H1. Better Internet banking system quality such as the transaction speed, ease of use, convenience, accessibility, cost/benefit, user empowerment, security, and privacy is likely to have a positive impact on customer satisfaction [30, 34, 46, 54].

H2. The difficulties of using technology due to customer characteristics such as age, education, income, disability and emotion are likely to moderate the relationship between Internet Banking system quality and customer satisfaction [17, 41, 46, and 55].

H3. The limitations of accessibility to technology due to customer system constraints such as availability of software, user's system processing speed and Internet bandwidth are likely to moderate the relationship between Internet banking quality and customer satisfaction [55].

H4. Customer satisfaction resulting from the usage of Internet banking is likely to have a positive impact on customer loyalty [13, 34].

6. CONCLUSION

This is a research in progress paper, providing an overview of Internet banking, addressing a number of issues identified from literature. Literature review discussed above address Internet banking as well as online banking customer issues. The hypothesis presented above will be empirically tested to identify customer satisfaction and loyalty in internet banking. Research will be guided by the conceptual model presented above as figure 2.

REFERENCES

1. Adams, L. S. & Martz, D. J. 2002, 'Developments in cyberbanking', *Business Lawyer*, vol. 57, no. 3, pp. 1257-364.
2. Ashford, W. 2004, 'Internet Banking Booms', in *Africa News Service*.
3. Australasian Business Intelligence 2002a, 'Internet banking: Australia enthusiastic but wary', *Australasian Business Intelligence*.
4. ---- 2002b, 'Macquarie gives up on account aggregation', *Australasian Business Intelligence*.
5. ---- 2002c, 'Wrapping up aggregation accounts', *Australasian Business Intelligence*, March 31, 2002.
6. Australian Bankers' Association 1999, Banking on the Internet - a guide to personal Internet banking services, <http://www.bankers.asn.au/ABA/pdf/ibpart1.pdf>, (Accessed 3 June 2004)
7. Barczak, G., Ellen, P. S. & Pilling, B. K. 1997, 'Developing typologies of consumer motives for use technologically based banking services', *Journal of Business Research*, vol. 38, no. 2, pp. 131-40.
8. Black, N. J., Lockett, A., Winklhofer, H. & Ennew, C. 2001, 'The adoption of Internet financial services: a qualitative study', *International Journal of Retail & Distribution Management*, vol. 29, no. 8, pp. 390-8.
9. BPAY 2002, BPAY Certifies New OFX Server Product from TPG, News Section 23 August 2002, [Web], http://www.bpay.com.au/about/news_detail.asp?id=7, (Accessed 8 July 2004)
10. Chang, Y. T. 2003, Dynamics of Banking Technology Adoption: An Application to Internet Banking, paper presented to Royal Economic Society Annual Conference 2003 41 / Royal Economic Society

(RePEc:ecj:ac2003:41).

11. Cunningham, R. M. 1956, 'Brand loyalty - What, Where, How much?' *Harvard Business Review*, vol. 34, no. 1, pp. 116-28.
12. Daniel, E. 1999, Provision of Electronic Banking in the UK and the Republic of Ireland, paper presented to International Journal of Bank Marketing, vol. 17, pp. 72-82.
13. Ford, W. S. Z. 1998, *Communicating with Customers*, Hampton Press. Inc., Cresskill, New Jersey.
14. Furst, K., Lang, W. W. & Nolle, D. E. 2000, Internet banking: Developments and Prospects, <http://www.occ.treas.gov/ftp/workpaper/wp2000-9.pdf>, (Accessed 2 May 2004)
15. Gerrard, P. & Cunningham, J. B. 2003, 'The diffusion of Internet banking among Singapore consumers', *International Journal of Bank Marketing*, vol. 21, no. 1, pp. 16-28.
16. Guest, L. 1944, 'A study of brand loyalty', *Journal of Applied Psychology*, vol. 28, pp. 16-27.
17. Ho, S. C., Chan, C. F. & Hsu, D. I. 1990, 'New Banking Technology in Hong Kong', *International Journal of Bank Marketing*, vol. 8, no. 6, pp. 9-15.
18. Howcroft, J. F., Hamilton, R. & Hewer, P. 2002, 'Consumer attitude and the usage and adoption of home-banking in the United Kingdom', *International Journal of Bank Marketing*, vol. 20, no. 3, pp. 111-21.
19. Jacoby, J. & Kyner, D. 1973, 'Brand loyalty vs. repeat purchasing behavior', *Journal of Marketing Research*, vol. 10, no. 1, pp. 1-9.
20. Jain, A. K., Pinson, C. & Malhotra, N. K. 1987, 'Customer loyalty as a construct in marketing of banking services', *International Journal of Bank Marketing*, vol. 5, pp. 49-72.
21. Joseph, M. & Stone, G. 2003, 'An empirical evaluation of US bank customer perceptions of the impact of technology on service delivery in the banking sector', *International Journal of Retail & Distribution Management*, vol. 31, no. 4, pp. 190-202.
22. Kahn, B. E., Kalwani, M. U. & Morrison, D. G. 1986, 'Measuring variety seeking and reinforcement behaviors using panel data', *Journal of Marketing Research*, vol. 23, pp. 89-100.
23. Karjaluoto, H., Mattila, M. & Pento, T. 2002, 'Factors underlying attitude formation towards online banking in Finland', *International Journal of Bank Marketing*, vol. 20, no. 6, pp. 261-72.
24. KPMG 1998, *Internet Banking Development, Operations and Maintenance*, KPMG Consulting Group, An Industry White Page.
25. Lee, D. 1999, Internet Banking and E-Commerce, www.authnet.com, (Accessed 4 May 2004)
26. Lee, E. J. & Lee, J. 2001, 'Consumer Adoption of Internet Banking: Need Based And/Or Skill Based', *Marketing Management Journal*, no. Spring 2001.
27. Leong, S. K., Srikanthan, T. & Hura, G. S. 1998, 'An Internet application for on-line banking', *Computer Communications*, vol. 20, pp. 1534-40.
28. Li, F. 2001, 'The internet and the deconstruction of the integrated banking model', *British Journal of Management*, vol. 12, no. 4, pp. 307-9.
29. Liao, S., Shao, Y. P., Wang, H. & Chen, A. 1999, 'The Adoption of Virtual Banking: an Empirical Study', *International Journal of Information Management*, vol. 19, no. 1, pp. 63-74.
30. Liao, Z. & Cheung, M. T. 2002, 'Internet-based e-banking and customer attitudes: an empirical study', *Information & Management*, vol. 39, pp. 283-95.
31. Mattila, M. 2003, 'Factors Affecting The Adoption Of Mobile Banking Services', *Journal of Internet Banking and Commerce*, vol. 8, no. 1.
32. Mattila, M., Karjaluoto, H. & Pento, T. 2003, 'A review of brand-loyalty measures in marketing', *Tijdschrift voor Economie en Management*, vol. 41, no. 4, p. 507(33).
33. Moullakis, J. 2004, 'Internet banking triples in three years.' *Australasian Business Intelligence*, Jan 26, 2004.
34. Moutinho, L. & Smith, A. 2000, 'Modelling bank customers' satisfaction through mediation of attitudes towards human and automated banking', *International Journal of Bank Marketing*, vol. 21, no. 1, pp. 5-15.
35. Mukherjee, A. & Nath, P. 2003, 'A model of trust in online relationship banking', *International Journal of Bank Marketing*, vol. 21, no. 1, pp. 5-15.
36. Ostroki, P. L., O'Brien, T. V. & Gordon, G. L. 1993, 'Service quality and customer loyalty in the commercial airline industry', *Journal of Travel Research*, vol. 33, no. 2, pp. 16-24.
37. Patricio, L., Fisk, R. P. & Cunha, J. F. 2003, 'Improving satisfaction with bank service offerings: measuring the contribution of each delivery channel', *Managing Service Quality*, vol. 13, no. 6, p. 471(82).
38. Pessemier, E. 1959, 'A new way to determine buying decision', *Journal of Marketing*, vol. 24, no. 2, pp. 41-6.
39. Plaskitt, S. 2002, 'Online banking numbers surge', *Australasian Business Intelligence*, July 24, 2002.
40. Polatoglu, V. N. & Ekin, S. 2001, 'An empirical investigation of the Turkish consumers' acceptance of Internet banking services', *International Journal of Bank Marketing*, vol. 19, no. 4, pp. 156-65.
41. Prendergast, G. P. 1993, 'Self Service Technologies in Retail Banking: Current and Expected Adoption Patterns', *International Journal of Bank Marketing*, vol. 11, no. 7, pp. 29-35.
42. Rotchanakitumnuai, S. & Speece, M. 2003, 'Barriers to Internet banking adoption: a qualitative study among corporate customers in Thailand', *International Journal of Bank Marketing*, vol. 21, no. 6/7, pp. 312-23.
43. Sathye, M. 1999, 'Adoption of Internet Banking by Australian Consumers: An Empirical Investigation', *International Journal of Bank Marketing*,

vol. 17, no. 7, pp. 324-34.

44. Sohail, M. S. & Shanmugham, B. 2003, 'E-Banking and Customer Preferences in Malaysia: An Empirical Investigation', *Information Sciences*, vol. 150, no. 3-4, pp. 207-17.
45. Suganthi, Balachandher & Balachandran 2001, 'Internet Banking Patronage: An Empirical Investigation of Malaysia', *Journal of Internet Banking and Commerce*, vol. 6, no. 1.
46. Tan, A., Beaumont, N. & Freeman, S. 1999, *Demographics and Perceptions as Determinants of Consumers' Use of Electronic Banking*, Working Paper Series, Department of Management, Monash University.
47. Turban, E., Lee, J., King, D. & Chung, H. M. 2000, *Electronic Commerce: A Managerial Perspective*, Prentice Hall, Upper Saddle River, New York.
48. Vatanasombut, B. 2001, Factors Affecting Retention of Customers Who Are Users of Computerized Application on the Internet: the Case of Online Banking, PhD thesis, Claremont University.
49. Vrechopoulos, A., G., S. & Doukidis, G. 2000, 'The Adoption of Internet Shopping by Electronic Retail Consumers in Greece: Some Preliminary Findings', *Journal of Internet Banking and Commerce*, vol. 5, no. 2.
50. Wall, S. 2001, 'Internet banking explosion', *Australian Banking & Finance*, vol. 10, no. 12, p. 1.
51. Wang, Y., Wang, Y., Lin, H. & Tang, T. 2003, 'Determinants of user acceptance of Internet banking: an empirical study', *International Journal of Service Industry Management*, vol. 14, no. 5, pp. 501-19.
52. Wright, A. & Ralston, D. 2002, 'The lagging development of Small Business Internet Banking in Australia', *Journal of Small Business Management*, vol. 40, no. 1, pp. 51-8.
53. Wu, C. H., Cheng, F. & Lin, H. 2004, 'Web site Usability Evaluation of Internet Banking in Taiwan', *Journal of Internet Banking and Commerce*, vol. 9, no. 1.
54. Yen, H. J. R. & Gwinner, K. P. 2003, 'Internet retail customer loyalty: the mediating role of relational benefits', *International Journal of Service Industry Management*, vol. 14, no. 5, pp. 483-500.
55. Zhang, P. & Li, N. 2004, 'An Assessment of Human-Computer Interaction Research in Management Information Systems: Topics And Methods', *Computers in Human Behavior*, vol. 20, no. 2, p. 125.
56. Singh, M., 2003, 'Managing a Secure E-Business Environment', The 2003 International Business Information Management Conference', December 16 –18, Cairo, Egypt, CD ROM.
57. Singh, M. (2000), 'Electronic Commerce in Australia: Opportunities and Factors Critical for Success', proceedings of the 1st World Congress on the Management of Electronic Commerce (CD ROM), January 19 –21, Hamilton, Ontario, Canada.
58. Chaffey, D., 2004, *E-Business and E-Commerce Management Second Edition*, Prentice Hall, UK.
59. Turban, E., King, D., Lee, J. and Viehland, D., 2004, *Electronic Commerce A managerial Perspective*, Prentice Hall, New Jersey

Internet Shopping Behavior among Singaporeans: Gender and Educational Issues

Hui Tak Kee, David Wan

School of Business, National University of Singapore, Singapore 117592
{bizhutik, bizwantw}@nus.edu.sg

ABSTRACT

The key objective of this study is to examine the influence of gender and educational level on attitudes toward online shopping. Based on a sample of 140 respondents, it was found that there is a general consensus over the convenience of the medium for information search and making purchase. Singaporeans are in general concerned with Internet privacy and security. Gender differences have a significant effect on the way respondents feel comfortable with not having the 'real' shopping experience. Both male and female respondents feel that it is easy to conduct online shopping and they generally encounter minimal problem. Finally, the perception of the Internet providing better prices and more cost savings increases with the level of education.

Keywords: Internet shopping, Online shopping, Net shoppers

1. INTRODUCTION

With the rapid advancement of new technology, all commercial activities including the retailing are affected in one way or another (Shaw, Gardner and Thomas, 1997). These changes not only create risks but also opportunities. The revolution is shaping the business world and at the same time pushing the US toward an information-based economy. In fact, the rate of technological change is so rapid that it will affect almost every aspect of how businesses are conducted. Singapore, being a small dot in the world map, is not spared.

Phau & Sui (2000) estimated that the population of Singaporean Internet shoppers to be 460,000 individuals in year 2000. This is a substantial figure given that the total population of Singapore is around four million people. The figures simply indicate that there are abundant opportunities for marketing managers who want to be a part of this growing market.

As such, understanding the attitudes, perceptions and behavior of the Singaporean Internet shopper allows the marketing professionals to efficiently target their marketing efforts. Furthermore, it also allows the marketing managers to better understand the concerns/fears of the Singaporean Internet shopper and subsequently work to address their demands and expectations. This will encourage more shoppers to shop over the Internet. This study aims to investigate the effects of gender and educational level on the attitudes of Singaporeans towards Internet shopping.

2. LITERATURE REVIEW

Before the age of the Internet, there have been many technological advances that influenced the operations of business transactions. The most recent technological advancement is the onset of Internet technology that

has transformed the entire business market place (Shaw, Gardner and Thomas 1997). Schwartz (1997) compared the way the Internet has changed the rules of business to the way automatic teller machines have transformed the customer's banking experience. The basis of comparison is that the customers are more comfortable doing things themselves with the ease, speed and convenience that Internet technology provides. In addition, the purchase of software over the Internet, sends the product directly into the buyer's computer and removes the olden day hassle of going to a shop, purchase the software and coming home to install it into the computer.

One of the clear advantages of shopping over the Internet is the fact that an online store does not incur the costs that a traditional store would incur. Thus costs like labor, fittings, rent and utilities are no longer relevant to an online store. As a result, products and services sold online can be priced at a cheaper rate (Lawrence, Corbitt, Tidwell, Fisher, Lawrence 1998).

There are several factors that are driving the progress of e-commerce, one of them being the increasing number of experienced users. When studying Internet shopping behavior, the familiarity of each user towards using the Internet has to be taken into consideration. Factors such as time spent on the Internet, the usage of Internet at work and the level of education all affect an individual's exposure and experience to Internet usage.

The Internet is also increasingly being used as a Customer Relationship Management (CRM) tool. CRM is not just another software or new technology but rather it deals with a more fundamental issue – unique customer needs (Geller 2002). Technology such as database management systems, using the Internet as a medium, has allowed marketers to pinpoint the individual preferences of each consumer and target

their marketing efforts in a way that appeals to the intended recipients and excludes the unintended ones.

Similar to any other innovation, companies that decide to use the Internet for marketing their offerings will need to meet or exceed the expectations of their customers to increase the likelihood of success. To this end, there are some important problems concerning the Internet's perceived security, competitiveness, impersonality and convenience that may affect the success of the companies marketing their products or services on-line.

One of the issues that consumer behaviorists have studied is the effect of time pressure on customer purchases. Research has shown that the likelihood of deferring the purchase is correlated to the effects of the pressure to make a decision in a given time as well as the type of product concerned (Dhar & Nowlis 1999). This study therefore aims to find out if time pressure affects the Internet shopper in the same way, given that while shopping over the Internet is 24hrs and in the comfort of the shopper's home environment.

Bell and Bucklin (1999) found that purchase decisions are based on the "Internal Reference Points", mainly past experience with a particular brand and product familiarity. Srivastava and Luire (2001) found that when search costs are low, the number of stores searched increases. On the other hand, if search costs are high, the consumers tend to accept the price-matching signal at face value and search less in the presence of a refund. It is thus appropriate to find out to what extent Internet shoppers value the low search costs and subsequently how that affects their buying behavior online.

One of the principle fears that consumers are facing when they are buying online concerns security and privacy (Bush, Bush and Harris 1998). The main fears relate to the possibility of the transaction intercepted by someone else (and therefore the personal particulars can be manipulated by a third party) as well as bogus or dishonest online vendor who fails to deliver the product and absconds or abuses the credit card particulars.

The effect of gender on Internet shopping is also a factor that is worth studying. Females have always been associated with a higher propensity to shop but it is the males that are found to be more involved with computer technology (Costa 1994). Internet shopping encompasses both shopping and technology and this study therefore aim to ascertain if there is a difference in attitudes between males and females towards this area. Alreck and Settle (2002) conducted a study in the U.S. and found no significant difference between the attitudes of both the male and female towards online shopping.

Another important demographic factor that is commonly used in predicting consumer attitudes is educational level. Research on this factor towards online shopping has not been particularly developed. This is probably due to the fact that the majority of the previous studies drew samples from website based surveys, thus eliciting self motivated responses from a demographically homogenous sample. In one particular census carried out in Helsinki, it was found that the higher a person's level of education, the more likely he/she uses the Internet for shopping (Statistics Finland 2002).

3. RESEARCH METHODOLOGY

In order to collect a random sample that would likely represent the population of Internet shoppers in Singapore, the island was classified into north, east, west and central areas. Four Mass Rapid Transit (MRT) train stations were randomly selected as target areas to conduct our survey. They are Yishun (north), Tampines (east), Boon Lay (west), and Raffles Place (central). A total of 140 surveys were collected from these four locations and the number collected from each location was equally distributed (35 surveys per MRT station). The survey was conducted at the main door of each MRT station and every 20th person was approached to participate in the survey. The collection of data was done over a period of four days (one station per day) between 1700-1900hrs. The reason for selecting this time frame was solely because of the higher pedestrian traffic during this period.

Respondents were required to complete a short (3 minutes) questionnaire that solicited their views on different aspects of Internet shopping, together with basic demographic information. The survey instrument consists of two sections. The first section comprises of twenty-eight 5-point Likert statements and the second section includes the demographic background. The construct of the questionnaire was designed to determine gender and educational level influences on a person's attitudes toward online purchase.

There are altogether 28 questions (attributes). Factor analysis was first conducted on these attributes to reduce their number to a more manageable size. Principal component analysis was used to extract the related factors first before the varimax rotation was used to further rotate for easy interpretations. Five factors were extracted. The factor scores were then used for analysis against the variables gender and educational levels in a 2-way ANOVA.

4. ANALYSIS AND FINDINGS

Of the 140 surveys that were conducted, 71 of them were males and 61 were females. Educational level was classified into 3 categories - Others, Intermediate and Upper. "Others" consists of those respondents with

up to secondary school level. The “Intermediate” category includes respondents having education up to junior college or diploma. Respondents who have a degree (or above) are classified under the “Upper” category.

There were 24 males (33.8%) who fell under the “Others” category, 35 (49.3%) fell under the “Intermediate” category and 10 (14.1%) in the “Upper” category. Six out of ten female respondents fell under the “Intermediate” category. There were only 12 (17.4%) and 14 (20.3%) who fell into the “Others” and “Upper” category respectively.

Of the 140 respondents, only 18 (12.9%) do not have Internet connection. 33 of the Internet users have bought online previously (14 male and 18 female). It was also found that majority of the male purchasers bought either software or multimedia products online. Online purchases of books and tickets were more prevalent amongst the females.

Results of Factor Analysis

To reduce the 28 questions to a more manageable size, factor analysis was used to identify the key factors included in this study. The number of factors is obtained from a scree plot which shows that only 5 factors have eigenvalue of 1 or higher. The total variance explained by these five factors is 72.5 percent.

Factor 1 comprises of five questions which are related to the issue of convenience in using the Internet for shopping and so it is named as “convenience” and its variance is 16.26 percent. Factor 2 comprises of five questions relating to “security concerns” such as “confident in giving out personal particulars” and “technology reliability” and the variance of this factor is 15.96 percent. Factor 3 involves those questions that deal with “personality” (variance = 15.78 percent). Factor 4 comprises of five questions which are concerned with whether it is easy to use the Internet for shopping. The factor is named as “user experience” and its variance is 14.25 percent. Finally, the last factor is related to “pricing” (variance = 10.26 percent).

Results of the 2-way ANOVA

The total scores of five components derived from factor analysis were tabulated against gender and educational levels. The total scores with respect to the male and female respondents are very close except for the “personality” factor where the deviation is obvious. However, for the education category, the deviations among different educational levels are more prevalent.

In this study, we would like to examine the respective role of gender and education on each factor attribute. The 2-way ANOVA technique is used for this purpose. Overall, there exist no interaction effect between

gender and educational level for all the five factors. The p-values vary from 0.16 to 0.94 (see Table 1). Therefore, this will warrant further analysis into each component. For the convenience factor, one can easily detect that there are no main effects in both gender and educational level ($p\text{-value} > 0.05$). In other words, there exists no difference between the male and female respondents with respect to their perception on the convenience factor. Both groups agree that it is very convenient to shop using the Internet and this can be carried out anytime they want. The same argument applies to the different educational achievements. Regardless of their educational levels, respondents feel that it is very convenient to do their shopping via the Internet.

For security concerns, there exists no main effect for gender ($p\text{-value} > 0.05$) but there exists a main effect for educational level ($p\text{-value} < 0.01$). The mean score for the higher educational group is 16.75 as opposed to 12.17 for the lower educational group. In other words, the higher educational group seems to have more confidence in giving out their particulars over the Internet, as compared to the lower educational groups.

For the personality issue, main effect exists in both cases (both $p\text{-value} < 0.01$). The male respondents (mean = 20.54) are more at ease with online shopping than their female counterparts (mean = 15.54). They believe that they are less likely to be hassled by the sale promoters to make purchases. They also do not mind looking up the information themselves. As for the different educational levels, it is again detected that the higher educational group (mean = 21.52) is more at ease than the lowest educational group (mean = 15.86).

For the user experience factor, no main effects are found for both gender ($p\text{-value} > 0.05$) and educational level ($p\text{-value} > 0.05$). It simply implies that both male and female respondents feel that it is easy to conduct online shopping and they generally encounter minimal problem. Similarly, different educational groups feel that they experience few problems in using the Internet technology.

Lastly, for the prices issue, the main effect exists marginally for gender ($p\text{-value} = 0.08$). However, if the significant level is assumed to be 0.05, then there exists no main effect. In other words, both male and female respondents feel that they can get a better deal by shopping online. It also allows them to compare the prices before they make the deal. On the other hand, perceptual differences exist for different educational groups. Those with university degree or above tend to agree more on the proposition that the internet shopping website offers competitive prices and also allows them to look for the best price before they purchase (mean = 22.59).

Table 1 Results based on 2-way ANOVA

Source	Sum of squares	DF	Mean Square	F value	<i>p-value</i>
<i>Convenience</i>					
Gender	2.68	1	2.68	1.53	0.22
Educational level	5.28	2	2.64	1.51	0.22
G x EL	5.18	2	2.59	1.48	0.23
<i>Security concerns</i>					
Gender	2.96	1	2.96	0.86	0.36
Educational level	302.13	2	151.06	43.76	0.00
G x EL	13.05	2	6.52	1.89	0.16
<i>Personality</i>					
Gender	705.83	1	705.83	62.27	0.00
Educational level	331.32	2	165.66	14.62	0.00
G x EL	38.89	2	19.44	1.72	0.18
<i>User experience</i>					
Gender	2.41	1	2.41	0.12	0.73
Educational level	20.65	2	10.33	0.50	0.61
G x EL	51.27	2	25.63	1.25	0.29
<i>Prices</i>					
Gender	18.35	1	18.35	3.10	0.08
Educational level	260.47	2	130.23	21.97	0.00
G x EL	0.79	2	0.40	0.07	0.94

In sum, there exists no interaction effect between the gender and the educational level in all the issues examined. For gender, main effect exists in the “personality” issue only. For educational level, main effect exists in 3 out of 5 cases, namely, “security concerns”, “personality” and “prices”.

5. DISCUSSION AND CONCLUSION

The main objective of this study is to examine the influence of gender and educational level on attitudes toward buying or consider buying over the Internet. The high means for the convenience factor showed that there was a general consensus among the respondents that the Internet is a convenient medium for information search or making purchase, taking into the account that the Internet spans international boundaries and is available 24hrs a day. This was indicative of individuals with different educational levels and of both gender groups.

However, Singaporeans in this study registered low means for security concerns regardless of gender or educational level. This can be attributed to the fact that Internet privacy and transactional systems are still in its infancy stages locally. With the recent spate of Internet

crimes involving banks and tampering of private information, the concern of revealing personal information on the Internet is reflected here. In addition, better-educated respondents seemed to be less concerned about security issues. This could be explained by the fact that better-educated individuals spend more time on the Internet. Thus familiarity levels and trust towards Internet security are higher as a result. This opens an avenue for companies tapping the local Internet market to design alternative strategies for market segmentation based on educational levels.

The personality factor measured the extent to which individuals are comfortable with not having a ‘real’ shopping experience, where physical touch and enjoyment enhance the shopping process. Gender differences were shown to have a significant effect on this factor. Females indicated a dislike for not being able to savor a physically shopping experience online while males did not mind that much. This of course could be very product specific as women often purchase clothes; cosmetics while the males tend to purchase more on products like software and financial services. This might serve as a working platform for Internet companies wanting to target the female consumer market. The lack of physical touch could be

substituted with a more interactive website, providing more information, or even making use of new internet technology like virtual tasting and touching.

The study also found that higher educated individuals are less likely to be affected by the personality factor. This is especially so for those who are higher educated. This can probably be attributed to the fact that the higher educated individuals spend more of their time at the work place and have less time to conduct physical shopping. They might also possess better information and knowledge on products and services they intend to purchase, and thus feel safe about not having to have physical feel or personal contact in the buying process.

No significant effect was detected for user experience for both gender and educational groups. User experience measured the perception of usability and usefulness of the Internet in general and a high total mean score indicated that Singaporean users are generally satisfied with the functions provided by the Internet. The fact that this is consistent through gender and all educational levels could indicate high Internet penetration rate among Singaporeans across the board. This is reflected in our data set where 87.1% of the respondents have Internet connection and six in ten users use the Internet for product information search. With the high Internet literacy rate and satisfactory user experience, there is a vast pool of consumers for which Internet companies could work on to identify their niches.

The study also predicts that educational levels would have an effect on the perception of pricing on the Internet. The perception of the Internet providing better prices and more cost savings increased with the level of education. In this study, the characteristics of people who actually search the Internet for information might provide a clue. Most of the respondents who claim they use the Internet for information search belong to the immediate or upper educational levels. The fact that they use the Internet for information search gives more opportunities for price comparisons as compared to those who do not do so. It is also noteworthy that web based businesses are able to offer more attractive prices due to reductions in capital costs of a physical operating outlet.

REFERENCES

- [1] Alreck P., R. B. Settle., "Gender effects on Internet, catalogue and store shopping", *Journal of Database Marketing*, Vol. 9, No. 2, pp150-163, 2002.
- [2] Anonymous, "Survey: CE Internet shopping gaining in popularity", *TWICE*, Vol. 13, No. 26, pp12, New York, 1998.
- [3] Bell R. D., R. E. Bucklin, "The role of internal reference points in the category purchase decision", *Journal of Consumer Research*, Vol. 26, No. 2, pp128-144, 1999.
- [4] Bush, A., V. Bush., S. Harris, "Advertiser Perceptions of the Internet as a Marketing Communications Tool", *Journal of Advertising Research*, Vol. 38, No. 2, pp17-27, 1998.
- [5] Costa, J. A., *Gender Issues in Consumer Behavior*, Sage, 1994.
- [6] Dhar R., S. M. Nowlis, "The effect of time pressure on consumer choice deferral", *Journal of Consumer Research*, Vol. 25, No. 4, pp369-385, 1999.
- [7] Geller L. K., "CRM: What does it mean?", *Target Marketing*, Vol. 25, No. 8, pp23-25, 2002.
- [8] Lawrence, E., B. Corrbitt., A. Tidwell, J. Fisher, J. R. Lawrence, *Internet Commerce: Digital Models for Business*, Wiley., 1998.
- [9] Phau I., M. P. Sui, "Factors influencing the types of products and services purchased over the Internet", *Internet Research*; Vol. 10, No. 2, pp102-113, 2000.
- [10] Schwartz, E. I., *Webonomics: Nine Essential Principles for Growing your business on the World Wide Web*, Broadway Books, 1997.
- [11] Shaw, M., D. Gardner, H. Thomas, "Research Opportunities in Electronic Commerce", *Decision Support Systems*, Vol. 21, pp27-33, 1997.
- [12] Srivastava J., N. Lurie, "A consumer perspective on price-matching refund policies: Effect on price perceptions and search behavior", *Journal of Consumer Research*, Vol. 28, No. 2, pp 296-308, 2001.
- [13] Statistics Finland, "E-commerce", 2002.
Internet:
http://www.stat.fi/tk/yr/tietoyhteiskunta/verkkokauppa_en.html

Markov Chain-based Clustering Analysis of Customers and WebPages

Changshou Deng^{1,2}, Pie Zheng¹, Yanling Yang², Bingyan Zhao²

¹Institute of Systems Engineering, Tianjin University, Tianjin 300072, China

²Jiujiang University, Jiangxi 332005, China
dengtju@eyou.com

ABSTRACT

This paper focuses on users' behavior towards an EC website. A novel Markov Chain-based way combining the web log file information and the topology of an EC website is presented to rank a user's interest in a WebPage. Then a URL-USERID relevant matrix is set up, with URL taken as a row and USERID as column, and each element's value is the probability of a user to access a WebPage when time goes infinitely. The similarity of each column vector can be used to cluster customers, and relevant web pages can be found from the similarity of each row vector. The knowledge discovered by this dynamic model can be fairly helpful to the design and maintenance of a website, to provide personalized service, and can be used in an effective recommending system of an EC website etc.

Keywords: Markov chain, electronic commerce, customers clustering, WebPages clustering

1. INTRODUCTION

With the high development of Internet, Web-based applications have been warmly welcomed. For instance, Remote Education, Network Bank, and in particular, the Electronical Commerce(EC) based on Internet have been increasing quickly. When a customer is browsing an EC website, his behavior information will be collected by the Web Server automatically, and stored in a weblog file. On the other hand, the topology of website's WebPages indicates some relationship between WebPages. If knowledge can be discovered from the weblog and the structure of an EC website, it can do good to the design and maintenance of a website, and provide personalized service, and Customer Relationship Management (CRM) can be conducted effectively. In a word, it can lead an EC website to succeeding in today's competitive business market. Customers clustering and WebPages Clustering are one of the most important issues in discovering knowledge from the Web. They are the preconditions for most web-based applications.

Intuitively, when time goes infinitely, users' ultimate probability of browsing each WebPage is a constant, and it should reflect the worthiness of each page to the users. This hypothesis is proved in this paper, and this has potential use in web-based data mining. Firstly, a user's behavior is abstracted as a homogeneous Markov Chain dynamically, which then is proved to be ergodic. Secondly with the help of the ergodic Markov Chain, a novel way to evaluate a user's interest in a WebPage combining the weblog and the structure of an EC website is presented. Lastly, a URL-USERID relevant matrix is set up, where URL is taken as a row and USERID is taken as column, and each element's value is the probability of a user to access a WebPage when time goes infinitely. Relevant WebPages and similar customers can be clustered efficiently by using the URL-USERID matrix.

This paper is organized as follows: Section 2 presents the necessary mathematical knowledge. The model of a user's behavior is presented in Section 3.

Section 4 presents the way to cluster the WebPages and customers. Section 5 concludes this paper and discusses the direction of the future research.

2. MATHEMATICAL BACKGROUND

Definition 2.1 Suppose the state space $\{S=i_0, i_1, i_2, i_3, \dots\}$ of the random variables $\{x(t), t=0, 1, 2, \dots\}$ is discrete; for only time t and any state space $i_0, i_1, i_2, \dots, i_{t-1}, i_t, j$, the equation:

$P\{x(t+1)=j, x(t)=i, x(t-1)=i_{t-1}, \dots, x(0)=i_0\} = P\{x(t+1)=j | x(t)=i\}$ is always true, then $\{x(t), t=0, 1, 2, \dots\}$ is named a Markov chain. $P\{x(t+1)=j | x(t)=i\}$ (abbreviated to $p_{ij}(t)$) represents the probability of transition from the state i to j at the time t . If the transition probability $p_{ij}(t)$ is independent of t , i.e., for any $i, j \in S$ and any different time t_1, t_2 , $p_{ij}(t_1) = p_{ij}(t_2)$, then this Markov chain is called homogeneous.

Definition 2.2 $P=(p_{ij})_{n \times n}$ is one step transition probability matrix of this homogeneous Markov chain, where

- (1) $p_{ij} \in (0, 1) \quad 1 \leq i, j \leq n$
- (2) $\sum_{i=1}^n p_{ij} = 1 \quad j = 1, 2, 3, \dots, n$

Definition 2.3 Suppose $\{x(t), t=0, 1, 2, \dots\}$ is a homogeneous Markov Chain on the state space S , the probability $p_{ij}^{(k)} = P\{x(t+k)=j | x(t)=i\}, i, j \in S$ represents the system will transit from state i to state j after k th step which is called k -step probability. And the matrix $P=(p_{ij}^{(k)})_{n \times n}$ is named k -step transition matrix, where

- (1) $p_{ij}^{(k)} \in (0, 1) \quad 1 \leq i, j \leq n$
- (2) $\sum_{i=1}^n p_{ij}^{(k)} = 1 \quad j = 1, 2, 3, \dots, n$

3. MODELING A USER'S BROWSING BEHAVIOR

For a virtual user A, $R=\{r_1, r_2, \dots, r_n\}$ denotes the set of URLs of an EC website, and S is considered as the space state. At time t , the user A is browsing the URL r_i in the probability $p_i(t)$, and will leave for the URL r_j in probability $p_{ij}(t)$. The user's surfing action on the web can be abstracted as a homogeneous Markov Chain, if $p_{ij}(t)$ is independent of time t .

Theorem 1 Suppose $\{x(t), t=0, 1, 2, \dots\}$ is a homogeneous Markov Chain on the state space S , for $m \geq 1, n \geq 1$ and $k = m + n$, the following is true.

$$p_{ij}(k) = \sum_{r \in S} p_r(m) p_{rj}(n) \quad i, j \in S \quad (3-1)$$

Eq.(3-1) is the famous Chapman-Kolmogorov equation. This equation denotes that if a user wants to move from space i to space j by $(m+n)$ steps, he must move to any space r by m steps and then move to space j by n steps.

The matrix form of this theorem is $P_{m+n} = P_m \cdot P_n$. Generally $P_n = P_1^n$ can be got. From the equation, the n -step transition matrix of homogeneous Markov Chain can be calculated by one step transition Matrix.

Theorem 2 At time t , a Markov Chain's distribution vector $p(t)$ is totally determined by the initial distributed vector $p(0)$ and one step transition matrix, that is $p(t) = p(0) P^t$.

3.1 Analysis of user's interest in WebPages

How to evaluate the importance of a WebPage and a user's interest in the WebPage is crucial to web data mining. As the user's interest in a WebPage is concerned, it should at least include the following metrics:

Relevance: It's an intuitive metric, denoting how much the user cares about the WebPage. It has been used as a useful metric in today's most Web Search Engine.

Authority: The authority metric means how many WebPages refer to the WebPage r . Moreover, the Web resources referred by higher quality resources should be assigned with higher authority.

Hub: The hub metric means how many WebPages are pointed by the Web resource r . Moreover, the Web resources pointing to higher quality resources should be assigned with higher hub.

Novelty: It denotes that how much a WebPage differs from other WebPages.

3.2 User's Browsing Behavior

Suppose that a user is browsing the WebPage of an EC website, his behavior is as follows. At time t , he is browsing the WebPage r_i and at next time $t+1$, he may have one of the following options:

(1) Continue browsing the WebPage r_i

(2) Click one hyperlink in r_i to jump to a new URL r_j ;

(3) Press the "BACK" button in the browser so as to return to the last browsing WebPage;

(4) To enter a new URL of the website.

Towards each options above, the probability of a user's tendency is measured as $\alpha, \beta, \gamma, \delta$ respectively, which satisfies the condition $0 < \alpha, \beta, \gamma, \delta < 1, \alpha + \beta + \gamma + \delta = 1$

Definition 3.1 Web Graph

The linkage structure graph of the WebPages, $G=(V,E)$, is a directed graph, where V is the set of nodes ($|V|=n$), a node corresponds to a WebPage for instance, in R ; and E is the set of edges, $E=\{(w_i)|v_i, v_j \in V \text{ and WebPage } i \text{ points to } j \text{ through hyperlink}\}$.

Definition 3.2 Tendency Matrix

The Tendency Matrix of resources R is $U=(u_{ij})_{n \times n}$ and

$$u_{ij} = \begin{cases} \alpha & i = j \\ \beta & (v_i, v_j) \in E \\ \gamma & (v_j, v_i) \in E \\ \delta & \text{otherwise} \end{cases} \quad (3-2)$$

It is noted that the Tendency Matrix here has synthesized the four metrics ($\alpha, \beta, \gamma, \delta$) mentioned above (Relevance, Authority, Hub and Novelty). With the tendency matrix being normalized, one-step transition probability matrix in the Markov Chain for a user's surfing behavior on the EC Website can be got.

Definition 3.3 One Step Transition Matrix

One step transition matrix of a virtual user A on the resource R can be defined as $P=(p_{ij})_{n \times n}$, and each element of P is as follows:

$$p_{ij} = \frac{u_{ij}}{\sum_{j=1}^n u_{ij}} \quad i = 1, 2, \dots, n \quad (3-3)$$

The distribution vector $p(t)$ can be calculated by theorem 2 and definition 3.3. It's obvious that $\alpha, \beta, \gamma, \delta$ actually reflect the relative importance, in the user's opinion, of relevance, authority, hub and novelty metrics respectively. And thus α is called relevance parameter, β authority parameter, γ hub parameter and δ novelty parameter.

3.3 Stationary distribution of a Markov Chain

Definition 3.4 Suppose that S is the state space of one homogeneous Markov Chain $\{x(t), t=0, 1, 2, \dots\}$, if for any $i, j \in S$, there exists a constant μ_{ij} which is independent of i , making the following equation always true. then the Markov Chain is ergodic.

$$\lim_{t \rightarrow \infty} p_{ij}(t) = p_j$$

Theorem 3

A homogeneous Markov chain, on the state space $S=\{s_1,s_2,...,s_n\}$, is ergodic, if for any $s_i,s_j \in S$, there exists a positive integer k , the state can transit from s_i to s_j via the edge between s_i and s_j in a positive probability within k steps, that is

$$\lim_{k \rightarrow \infty} p_{ij}^{(k)} = p_j$$

where $p_j(j=1,2,...,n)$ is the unique solution to the following equation:

$$p_j = \sum_{i=1}^n p_i p_{ij} \quad (j = 1,2,3,...,n) \quad (3-4)$$

which satisfies the restricted (3-5) conditions.

$$p_j > 0 \quad (j = 1,2,..., n)$$

$$\sum_{j=1}^n p_j = 1 \quad (3-5)$$

Theorem 3 denotes that when a homogeneous Markov Chain is ergodic, the distribution vector $p(t)$ is only determined by one step transition matrix, independent of time t and of it's initial vector $p(0)$.

Definition 3.5 Positive Stochastic Matrix

Suppose $P=(p_{ij})_{n \times n}$ is an n -dimensional non-negative matrix, if it satisfies the following two conditions:①the sum of all elements in each row is 1 ②if there exists positive integer k ,making $P^k > 0$,then the matrix P is called positive stochastic matrix.

It is true without proof that each element in one-step matrix is positive and the sum of each row is 1, respectively. So one step matrix is positive stochastic matrix.

Theorem 4 A Markov Chain on the finite State Space $S=\{s_1,s_2,s_3,...,s_n\}$ is ergodic if and only if it's transition matrix is positive stochastic matrix.

In a word, with the help of formulas (3-4) and (3-5), the ultimate probability can be got easily. Suppose the user A is rational and experienced enough in an EC website, the probability of browsing the WebPages should be proportional to its worthiness. In real life, user usually spends more time with the WebPages he cares more. Firstly,since the user has no knowledge about the value of each WebPage, he browses WebPages randomly. The user becomes more and more experienced while more and more WebPages he browsed. So his judgment on the value of WebPages becomes more and more accurate. When time goes infinitely, the ultimate probability of browsing each WebPage should reflect the worthiness of each WebPage to the user accurately. From Theorem 3,the ultimate distribution vector $p(t)$ of a ergodic Markov chain is independent of the initial distribution vector $p(0)$, but totally determined by the one step

transition probability matrix. So given an EC Website WebPages R , the transition probability matrix can be set up through definition 3.3,and the ultimate distribution vector can be calculated based on Theorem 3. The ultimate distribution vector is user's interest in WebPages. The group of equations in Theorem 3 can be solved using "Gauss method".The ultimate constant probability of user's browsing the web can be used to discover the similar customers and relevant WebPages.

4. CUSTOMERS CLUSTERING AND WEBPAGES CLUSTERING

4.1 Data Preparation

An EC website and the weblog file will be used in knowledge discovering. The weblog file complies with W3C standard. It at least includes the following items:

- ①user's IP address ;②Access time;③Request method;
 - ④the URL of the accessed WebPage;⑤the transferred protocol;⑥return code;⑦the bytes transferred.
- Before discovering knowledge from weblog,data preparation is needed. It has two stages: data cleaning and transaction identification. Suppose L is a weblog,and l belongs to L .The items of l are,respectively, $l.ip,l.uid$ and $l.url$.The algorithm to find a transaction is as follows:
Step1: to partition weblog by the user's IP address,that is to determine each visitor's browsing set;
Step2:to calculate the times of each customer's browsing each WebPage in every set;
Step3: to sort the browsing set by IP address, and then the users' browsing set can be got.

4.1.1 Topology of an EC website

A website's topology is a directed graph, and customer's browsing pattern is its subgraph. The customers who have similar visiting subgraphs are similar. Suppose that an EC website's topology is as the following directed graph: $G=(N, N_p, E)$,where N is the set of nodes, $N_p(N_p=\{Node \in N, \{userid,p\}^n \mid n \geq 1\})$ stores the customers' USERID and the ultimate probability to browse each Webpage, which can be calculated by theorem 3. N is the attribute of node; E is the set of edges. From the set N of G , all the urls of the EC website can be got. And from the each node's userid and his ultimate probability to visit this WebPage can be found. According to this, A URL-USERID relevant matrix $M_{m \times n}$ can be set up as follows:

$$M_{m \times n} = \begin{bmatrix} p_{11}, p_{12}, L, p_{1n} \\ M \quad L \quad M \\ p_{m1}, p_{m2}, L, p_{mn} \end{bmatrix}$$

Analysis of the matrix can be made as follows.

- (1) Each element p_{ij} of this matrix presents the probability of the j th customer's to visit the i th URL when time goes infinitely.
- (2) The i th row vector represents that the probabilities

of the customers to visit the *i*th URL when time goes infinitely.

- (3) The *j*th column vector represents the probabilities of the *j*th user to visit each URL when time goes infinitely. And the other form of matrix $M_{m \times n}$ can be written as $M_i = (p_1, p_2, \dots, p_n)$.

At the same time, URL-USERID relevant matrix can be represented by row vectors, and each row vector is of a constant. So a conclusion can be drawn that the row vectors represent the structure of the EC website and also imply the common browsing pattern; the column vectors represent the types of customers and the personalization browsing pattern. The similarity of each row vector represents the relevant WebPage and the similarity of each column vector represents the similar customers. The similarity can be judged by the definition 4.1. The larger value of the angle of between vectors *a* and *b*, the more similar they will be. In practical use, a threshold can be set to determine the similarity. Definition 4.1 Suppose vectors $a = (a_1, \dots, a_n)$, and $b = (b_1, b_2, \dots, b_n)$, then, the cosine of the angle between vectors *a* and *b* is

$$\cos(a, b) = \frac{a \cdot b}{|a| \cdot |b|} = \frac{(a_1 \cdot b_1 + a_2 \cdot b_2 + \dots + a_n \cdot b_n)}{\sqrt{a_1^2 + a_2^2 + \dots + a_n^2} \cdot \sqrt{b_1^2 + b_2^2 + \dots + b_n^2}}$$

4.1.2 One-Step Matrix and Ultimate Probability

Four metrics described in Section 3 can be used to evaluate the interest of a user in a WebPage during given time *t*. Times of the user's visiting the WebPage represent the relevance of a user to a WebPage. The inverse of a url's Indegree (the number of hyperlinks pointed to the URL) represents the authority; the Outdegree (the number of hyperlinks pointing to the other URL) of the url represents the hub of a url; the novelty can be judged by the inverse of the Indegree of a url approximately. The detailed method is as following.

Firstly, the trend matrix $U = (u_{ij})_{n \times n}$ (at time *t*) is formed by the following

$$u_{ij} = \begin{cases} \alpha & i = j \\ \beta & (v_i, v_j) \in E \\ \gamma & (v_j, v_i) \in E \\ \delta & \text{otherwise} \end{cases}$$

Secondly, the following is defined

$$a = p_1 \times \frac{l_i^t \cdot \text{time}}{\sum_{i=1}^m l_i^t \cdot \text{time}} \quad b = p_2 \times \frac{\text{outdegree}(v_j)}{|E|}$$

$$c = p_3 \times \frac{\text{indegree}(v_i)}{|E|} \quad d = p_4 \times \frac{1}{\text{indegree}(v_i)}$$

Third, normalizations of *a, b, c, d* is as follows,

$$a = \frac{a}{\sqrt{a^2 + b^2 + c^2 + d^2}} \quad b = \frac{b}{\sqrt{a^2 + b^2 + c^2 + d^2}}$$

$$c = \frac{c}{\sqrt{a^2 + b^2 + c^2 + d^2}} \quad d = \frac{d}{\sqrt{a^2 + b^2 + c^2 + d^2}}$$

respectively.

let

$$\alpha = a^2 \quad \beta = b^2 \quad \gamma = c^2 \quad \delta = d^2$$

It can be noticed that

- (1) p_1, p_2, p_3 and $p_4 (p_1, p_2, p_3, p_4 > 0)$ are coefficients.
- (2) If the outdegree or indegree of a node is zero, then let the outdegree or indegree is 1.
- (3) It is obvious that *a, b, c, d* can be adapted by the coefficients p_1, p_2, p_3, p_4 .

4.2 Customer clustering and WebPages clustering

With the help of the above URL-USERID matrix, clustering can be done in an EC website.

4.2.1 Customers Clustering

As described above, the column vector (p_1, p_2, \dots, p_n) of relevant matrix $M_{m \times n}$ is the personalized browsing pattern. The customers who have similar browsing patterns can be clustered as similar customers. The cosine of the angle between the two column vectors can be calculated by definition 4.1 to cluster similar customers. In order to cluster the similar customers, a matrix *Q*, of which each element is the cosine of the angles between two column vectors, can be formed. It's obvious that the matrix *Q* is symmetrical, i.e. $Q = (q_{ij})_{n \times n} q_{ij} = q_{ji}, q_{ii} = 0$. The matrix *Q* is as follows:

$$Q = \begin{bmatrix} 0 & q_{12} & \dots & q_{1n} \\ q_{21} & 0 & \dots & q_{2n} \\ \dots & \dots & 0 & \dots \\ q_{n1} & q_{n2} & \dots & 0 \end{bmatrix}$$

A threshold is given by formula(4-1)

$$\theta = \frac{2 \times \sum_{i=1}^{n-1} \sum_{j=i+1}^n q_{ij}}{n \times (n-1)} \quad (4-1)$$

The threshold is the average of all elements in matrix *Q*, and it can be adapted to satisfy the requirements in practical usage. For any $q_{ij} \in Q (1 \leq i \leq n, i < j \leq n)$, if $q_{ij} > \theta$, then the *i*th customer and the *j*th customer can be clustered into the same group. The more the threshold is, the more accurate the clustering can be.

All in all, any two customers can be clustered if the similarity of their ultimate probability to each WebPage vector satisfies the given threshold. In our daily life, as

time goes by, the similar customers will come into being after their long browsing and transaction. The method proposed in this paper can track this procedure and grasp the trend of this procedure.

complexity and time complexity are $O(n)$. When similar customers are to be discovered, it is only needed to evaluate the similarity of the two n -dimension vectors. Both space complexity and time complexity are $O(n)$.

4.2.2 WebPages clustering

As has been said that the row vectors represents what all the customers have browsed each WebPage of an EC website. If all the customers have the similar experience in some WebPages, then these WebPages will be thought as relevant WebPages. A matrix Q' is introduced, elements are the cosines of the angles of between row vectors. For ease usage, let $q'_{ij} = 0$ (if $i=j$). The matrix Q' is as follows:

$$Q' = \begin{bmatrix} 0 & q'_{12} & \dots & q'_{1m} \\ q'_{21} & 0 & \dots & q'_{2m} \\ \dots & \dots & \dots & \dots \\ q'_{m1} & q'_{m2} & \dots & 0 \end{bmatrix}$$

The threshold can also be calculated by formula (4-1) and adapted according to the requirements. For any $q'_{ij} \in Q'$ ($1 \leq i \leq m, i < j \leq m$), if $q'_{ij} > \theta$, then the i th URL and the j th URL ($j \in [i, m]$) can be considered as one group.

In a word, if all the customers browse a given WebPage, their ultimate probability of the WebPage is a constant. The row vectors of URL-USERID are also constants. For two different WebPages, if all the customers' probabilities of browsing them are the same or similar, it is reasonable that they are relevant WebPages.

4.3 Algorithm analysis

The method proposed in this paper to cluster relevant WebPages and similar customers is based on Markov Chain, which combines the weblog file and website topology. The way to evaluate a user's interest in a WebPage has integrated four metrics: Relevance, authority, hub and novelty. The linear equations can be used to calculate user's interest in WebPages. When WebPages are clustered, we only need to evaluate the similarity of the two n -dimension vectors. Both space

5. CONCLUSION

A Markov Chain-based dynamic model is put forward to investigate the customers' browsing behavior towards an EC website. As is well known that the similar customers will come into being after their long browsing and transacting with a EC website as time goes by. It is assumed that the ultimate probability of browsing each WebPage reflects the worthiness of each WebPage when time goes infinitely, and, thus, this probability is a constant. A URL-USERID relevant matrix is set up to cluster customers and WebPages. This novel method can be intergrated in an EC recommending system and in personalized service system. That will be our further research interest.

REFERENCES

[1] S. Brin, L. Page, "The anatomy of a large-scale hypertextual Web search engine", Computer Networks and ISDN Systems, Vol.30, No. 7: pp107-117,1998.

[2] MS.Chen, JS.Park, "Data Mining for Path Traversal Patterns in A Web Environment", proceeding of the 16th Int'l Conf on Distributed Computing Systems, pp385~392, HongKong, May, 1996.

[3] O.R.Zaiane, M.Xin, J.Han, "Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs", proceeding of Advances in Digital Libraries Conf, pp19~29, Santa Barbara, CA, 1998.

[4] Qin Bao Song, Junyi Sheng, "An Efficient and Multi-Purpose Algorithm for Mining Weblog", Computer Research and Development, Vol.38, No.3, pp328~332, 2001.

[5] Rong xing Wang, Stochastic Process, Xi'an Jiao Tong Univer sity press, 1987.

[6] Dell Zhang, Yisheng Dong, "An Efficient Algorithm to Rank Web Resources", proceeding of the 9th www conference on THE WEB: THE NEXT GENERATION. Amsterdam, May 15, 2000.

Modeling Competitive Diffusion Process in Japanese ADSL Service Market

Nozomi Nakajima¹, Jian Chen²

¹ Graduate School of Economics, Osaka University, Osaka, 5600043, Japan

² Integrated Marketing Center, Dentsu Inc., Tokyo, 1057001, Japan
nakajima@econ.osaka-u.ac.jp, chenjian@dentsu.co.jp

ABSTRACT

This study models innovation diffusion process with brand-level competition, focusing on the competition among ADSL service providers in Japan. For a particular brand (service provider), the diffusion process is assumed to be influenced by three forces: (1) the external influence through mass media, (2) the internal influence of the communication with the brand adopters, and (3) the influence due to the market growth of the product category. Through the assumption that the internal influence of each brand in a product category has the identical structure, the proposed brand-level model can be summed up to the Bass model of the product category, and solved with a closed-form expression. Applying the model to the diffusion of ADSL market in Japan, the empirical results reveal that the proposed model describes the brand-level diffusion patterns very well.

Keywords: ADSL service, diffusion process, competition, Bass model

1. INTRODUCTION

Following the success of the Bass model (1969), marketing researchers have developed different types of diffusion models to address various issues concerning the sales growth of new products (Krishnan, Bass, and Kumar 2000). These issues include analyzing the role of marketing-mix variables in the diffusion process, developing multi-product interaction diffusion models, investigating the innovation diffusion process in multi-markets, and so forth. However, most of the models focus only on category-level diffusion problems. For companies such as Intel and Microsoft that hold near-monopoly positions in their respective markets, the category-level sales growth is of primary concern. But in industries such as minivans and cellular-phones with severe competition, brand managers are likely to pay attention to understanding the sales growth at the brand-level.

The diffusion literature recognizes that all categories evolve from a centralized diffusion process, initially dominated by a monopolist, into a decentralized process having many competitors (Rogers, 1995). Cross-brand or competitive interpersonal influence was first investigated in the marketing literature by Peterson and Mahajan (1978). They proposed a system of equations for different types of multiple product interactions including complement, substitute, and contingent. Since the 1990s, researchers have proposed several models to describe brand-level innovation diffusion process. Currently, researches on brand-level innovation diffusion are not sufficient and always concern a special problem. As pointed out by Chatterjee, Eliashberg, and Rao (2000), there are opportunities for empirical researchers to work on competitive diffusion model specifications across product and service categories and across competitive positions within categories.

The paper first reviews the past researches about the brand-level innovation diffusion. Then a competitive diffusion model is proposed and applied to the diffusion of ADSL subscribers in Japan, followed by some managerial implications and limitations.

2. BRAND-LEVEL DIFFUSION MODELS

Mahajan, Sharma, and Buzzell (1993) extended the Bass model to assess the impact of new-brand entry on market expansion of the incumbent brands as well as the whole market. Specifically, they introduced a new interaction term to capture the internal influence of one brand's growth through drawing adopters from other brands' potential adopters. The model was illustrated by applying it to the Polaroid and Kodak case in instant photography industry from 1976 to 1985. The model was estimated under the assumption that the intensity of internal influence is identical for all brands. Estimation results suggest that the model is appropriate for the Polaroid and Kodak case.

Parker and Gatignon (1994) developed a framework to provide a systematic analysis of alternative specifications for brand-level first purchase diffusion. They started with a general specification in which the brand-level adoption is a function of diffusion effects, price and advertising, and suggested four types of brand-level diffusion processes depending on the nature of interpersonal communication and brand-level competition. They have evaluated these specifications by estimating these four models on brand trial data in a new packaged goods category, using the NLS procedure. The results showed that no one specification appeared to dominate others across brands, and the effectiveness of the models appeared to be significantly different across brands.

Givon, Mahajan, and Muller (1997) provided an example of a competitive brand level diffusion model with a more specific focus than Parker and Gatignons' (1994) work. In a competitive extension to their earlier model (Givon, Mahajan, and Muller 1995), the authors investigated the implications of piracy in a duopolistic software market. The market dynamics are captured in a framework that extends the Bass model to incorporate brand switching as well as legal and illegal uses of the competing software brands. This modeling approach was used to analyze the diffusion of Lotus 1-2-3 versus Other Spreadsheets and WordPerfect versus Other Word Processors in the United Kingdom. The results showed that for Lotus 1-2-3 the users were likely to switch to Other Spreadsheets than vice versa, and for WordPerfect the users were likely to influence the users of Other Work Processors to switch than vice versa.

Krishnan, Bass, and Kumar (2000) proposed a model to analyze the impact of later entrant on the sales growth and diffusion speed for the whole category as well as for the incumbent brands. The authors supposed that two brands were presented in the market from the time of introduction. When the third brand was introduced, the market potential of the whole category would expand and the category market would start diffusing faster. Furthermore, the diffusion speed of the incumbent brands might be affected either positively or negatively depending on the model parameters. The model was applied to the cellular phone industry of six markets in the United States from 1983 to 1996. The results revealed that the effects of the third entry on incumbent brands and total category were different across markets.

Past researches on brand-level innovation diffusion have provided limited knowledge to understand both of the brand-level growth pattern and the relationships among brands. Researchers have made efforts to specify model parameters appropriate for special cases. All these models seem appropriate for the special cases in their works. However, the models cannot be solved in closed-form expressions, which made researchers employ discrete formulations to estimate continuous diffusion processes. This would result in the estimation bias and decrease the attractiveness of the models. Krishnan, Bass, and Kumar (2000) proposed a brand-level diffusion model for investigating the effects of a new entrant on diffusion speed and market potential for both the incumbent brands and the whole product category. Although their model has a closed-form solution, it did not incorporate the competitive effect among brands directly.

3. THE PROPOSED DIFFUSION MODEL

The concept underlying the Bass model is the assumption that the diffusion of an innovation is stimulated by two forces: (1) the external influence such as the mass media, and (2) the internal influence such as the communication between adopters and potential

adopters. For a particular brand in a product category, the brand-level diffusion may be affected by three forces: (1) the external influence through mass media, (2) the internal influence through adopter-nonadopter communication of the brand, and (3) the influence from the growth of product category (Nakajima, 1990). Hence, given that there are n brands in a product category, the diffusion process of the brand j ($j=1, \dots, n$) can be expressed as follows:

$$f_j(t) = (1 - F(t))(p_j + q_j F_j(t) + r_j F(t)) \quad (1)$$

where $F(t)$ and $F_j(t)$ are the cumulative adoption distribution for the product category and brand j , respectively. $f_j(t)$ is the p.d.f. of adoption for brand j . The term $p_j(1-F(t))$ captures the adoption due to external influence for brand j , and $q_j F_j(t)(1-F(t))$ captures the adoption due to internal influence of brand j . The parameters p_j and q_j represent external and internal coefficients, respectively, which are the same as those in the basic Bass model. The term $r_j F(t)(1-F(t))$ in the above equation represents the adoption due to the influence of the growth of the whole product category on brand j . For each brand, the influence from the maturity of total market may be different. This difference is captured by the parameter r_j .

It is common that the success of product category would influence the adoption of a particular brand in several ways, such as communicating the positive information about the product category, decreasing the uncertainty of the product, and so forth. Although brand-level communication would be the main force impacting the subsequent adoption of that brand, the effect of the cumulative adoption level of product category must be considered. Note that in equation (1), the adoption for brand j comes from the rest of the whole market potential, that is, $1-F(t)=1-\sum_i F_i(t)$. As Krishnan, Bass, and Kumar (2000) has pointed out, when consumers make decision to buy a new product, the question of "what brand to buy" would be only the secondary to the "whether or not to buy the product" question. In this case the category effect is likely dominant and the specification of market potential for each brand may not be suitable. If the category effect is not dominant, in contrast, the adoption for brand j can be expected to come only from $1-F_j(t)$ as in Mahajan, Sharma, and Buzzells' model (1993). In effect, for analyzing category-level diffusion, the Bass model (1969) which focuses on $f_j(t)/(1-F(t))$ should be used. On the contrary, for analyzing the brand-level diffusion, in which the category effect dominates the brand-level effect in consumer's buying behavior, $f_j(t)/(1-F(t))$ should be considered.

Summing up the both sides of equation (6) including all brands yields:

$$\sum_j f_j(t) = (1 - F(t))(P + \sum_j q_j F_j(t) + R F(t)) \quad (2)$$

where $P = \sum_j p_j$, $R = \sum_j r_j$. Furthermore, assuming that the intensity of internal communication effect for each brand in the product category is identical, that is, $q_j = Q$,

the above equation results in the category Bass model (Bass 1969):

$$f(t) = (1 - F(t))(P + (Q + R)F(t)), \quad f(t) = \sum f_j(t). \quad (3)$$

And for brand j ,

$$f_j(t) = (1 - F(t))(p_j + Qx_j + r_j F(t)) \quad (4)$$

Thus, the cumulative diffusion rate $F(t)$ can be easily solved as a function of time t :

$$F(t) = \frac{1 - C \cdot e^{-(P+Q+R)t}}{1 + (Q+R)/P \cdot C \cdot e^{-(P+Q+R)t}} \quad (5)$$

$$C = \frac{1 - F_0}{1 + (Q+R)F_0/P}, \quad F_0 = F(0). \quad (6)$$

Substituting $F(t)$ into the brand level equation (4) and then it can be solved to yield $F_j(t)$ as a function of time:

$$F_j(t) = \frac{Pr_j - p_j R}{QR} + \frac{r_j}{R} F(t) - K_j \left(F(t) + \frac{P}{Q+R} \right)^{Q(Q+R)} \quad (7)$$

$$K_j = \frac{r_j}{R} \left(F_0 + \frac{P}{Q+R} \right)^{Q(Q+R)} - \left(F_{j0} + \frac{p_j}{Q} - \frac{r_j P}{Q(Q+R)} \right) \times \left(F_0 + \frac{P}{Q+R} \right)^{Q(Q+R)}, \quad F_{j0} = F_j(0) \quad (8)$$

Unlike the model proposed by Krishnan, Bass, and Kumar (2000), the model proposed here incorporates the effects of category growth on adoption for a particular brand directly. The parameter r_j represents the intensity of influence of the growth of product category on brand j , and r_j may differ across brands. This turns to be important when considering the brand-level innovation diffusion issues. Furthermore, the brand-level model proposed here can be summed up to the Bass model (1969). This gives validity to the proposed model because the Bass model has a strong behavioral basis and has found excellent empirical support over a wide range of products (Mahajan, Muller, and Bass 1993). The assumption that $q_i = Q$ seems reasonable because the communication pattern may be the same for all brands in a product category. In addition, the closed-form expression of the proposed model has some other advantages, especially it enables researchers to employ the NLS estimation procedure.

4. APPLICATION: DIFFUSION OF ADSL SERVICE IN JAPAN

4.1 The ADSL Market in Japan

According to the Japan Ministry's statistics, by the end of September 2002, the number of ADSL subscribers in Japan totaled about 4.22 million, adding about 1.8 million or an increase of 77.5 percent as of a-half-year before, at the end of March 2002¹. This number breaks

down to 1.72 million subscribers of NTT East and NTT West, and 2.50 million subscribers of other service providers.

The ADSL market of Japan expanded rapidly with the entrance of large competitors. At the end of 1999, Tokyo Metallic Communications and eAccess began to launch internet connection service –ADSL for home users by offering data transfer speeds up to 640kbps. It cost subscribers more than 6,000 yen per month at that time. The ADSL market experienced slow-growth in the first year. But with the entry of NTT, the market began to take off. The ADSL had not been a viable choice for NTT, because NTT had decided to employ optical-fiber cables for access lines instead of copper-cable based technologies (ADSL). However, it would be projected to take more than ten years to complete the overall implementation. Therefore NTT decided to adopt ADSL during the construction of the optical-fiber networks. By the end of June 2001 when NTT had entered into the ADSL service market about half a year ahead of Yahoo!BB's new entry, the total number of cumulative subscribers increased up to 300 thousands. This number was about thirty times larger than that of a half-year before when NTT entered into the market. The market share of NTT was as high as 63 percent. In addition, not only the NTT, other companies such as Tokyo Metallic and eAccess also expanded their markets more rapidly than before.

With the entry of Yahoo!BB, the competition among companies became intense, especially in terms of price reduction. Before Yahoo!BB entered into the market, average monthly fees of the top-three companies (NTT, Tokyo Metallic Communications, and eAccess) were around 5,400 yen, which is approximately 20 percent reduction compared with that of the previous year. When Yahoo!BB offered monthly fee of 2,280 yen in August 2001, the large price reduction contributed to the rapid expansion of the market. The number of monthly subscribers increased from 110 thousands in July 2001 to 270 thousands in October 2001. Thereafter, the number of monthly subscribers remained around 330 thousands. From the beginning of 2002, the market share of NTT East and NTT West remained virtually unchanged about 40 percent. This means the market shares of NTT and the non-NTT camp were stabilized. However, the market shares among the non-NTT camp had been changed. According to the ministry's statistics, with the integration of Metallic Communications Corp. into the Softbank group, Yahoo!BB held about 21.6 percent of Japan's ADSL market. It appeared that Yahoo!BB and NTT began fierce competition for the top market share. With the development of ADSL technology, the maximum speed for data transmission has been changed from 640kbps to 26Mbps. Now, the difference of ADSL monthly fees among the companies is not large –from 3,500 to 4,000 yen. At present, the competition among companies is mainly based on service and quality but not price.

¹ To investigate the competition between the NTT group and the Others in the ADSL market, period from Jan. 2000 to Sep. 2002 is under consideration.

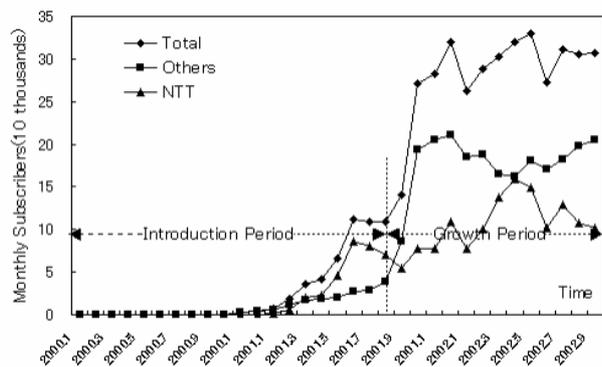


Figure 1 Monthly subscribers of ADSL in Japan

Figure 1 outlines the growth of monthly subscribers for the total market, the NTT group, and the Others. Since the data on the number of subscribers of each company could not be obtained easily, we used the data provided by the ministry's statistics to analyze the competitive relationship between the NTT group (the total of NTT East Corp. and NTT West Corp.) and the Others (the total of other companies) in this study.

4.2 Methodology

At the time when the analysis was conducted, the ADSL market in Japan was at the growth stage finishing up the introduction stage. With the entrance of large players like NTT and Yahoo!BB, the market grew up rapidly. In the concept of the proposed model, three forces – the external communication, the internal communication of a particular brand, and the growth of the total market would influence the diffusion of the particular ADSL service in a social system. Though both the NTT group and the Others benefited from the growth of the total market, the magnitude of the influence might be different. The influence of the total market on NTT would be large, since it was a popular brand among the people. On the other hand, the growth of the total market and the cumulative knowledge about the ADSL service made people consider the ADSL providers in a reasonable way. That is, the service offered by each provider did not differ significantly. Hence, the Others also benefited on some extent from the growth of the total market.

From August 2001, when Yahoo!BB entered into the market and announced to provide a low monthly fees of only 2,280 yen, the market demonstrated the aspects of the growth stage. The intense competition brought rapid price reduction and market expansion. Since Yahoo!BB was thought to have the advanced technology as well as low priced and high quality service, it would make some change in the competitive structure of the market. In this period, the influence from the growth of the market on the adoption of NTT service was expected to be getting similar with that of the Others. The proposed model is used to analyze the competitive relationship of ADSL service in Japan. The closed-form solution of the proposed model was used for estimation. Since the

characteristics of the ADSL market varied largely before and after Yahoo!BB's entry into the market, a two-period estimation was executed.

4.3 Empirical Results

Monthly cumulative data of ADSL subscribers for the NTT group and the Others were used for empirical analysis. The data were simultaneously fitted to the equation (7) ($j=1, 2$), employing the NLS procedure. The market potentials, however, must be separately estimated before conducting the NLS procedure. Since the ADSL market in Japan changed greatly after Yahoo!BB had entered into the market, the market potentials were also assumed to be changed. Therefore two different market potentials corresponding to before/after the Yahoo!BB's entry should be estimated. Although recent applications of diffusion models reported better forecasting results by using exogenous sources of information such as market surveys, secondary sources, management judgments, or other analytical models (Gatignon et al., 1989, Mahajan and Sharma 1986), the estimation of the market potential could be derived directly from the diffusion time-series data. In detail, applying the original Bass model to the category data for each periods, the market potentials of the introduction stage and the growth stage were estimated as 0.75 million and 5.55 million, respectively. We also assumed that p the innovative influence for the NTT group equal to zero in the estimation step. This is somewhat similar to the method used by Krishnan, Bass, and Kumar (2000).

Table 1 Estimation results of diffusion parameters

diffusion parameters	introduction stage (2000.1-2001.7)	growth stage (2001.8-2002.9)
Q	.3510 (.0741)	.1092 (.0408)
p_1	.0000034 (.00000093)	.0140 (.0014)
r_1	.0997 (.0242)	.0448* (.0212)
r_2	.2115 (.0323)	.0502 (.0184)

*: significant at 0.05; others: significant at 0.01

() : standard error

The estimation results are represented in Table 1. All the estimates are significant at 0.01 level except the one with asterisk mark. The results showed that the diffusion patterns for the NTT group and the Others are different. As for the Others, the three forces – the external influence, the internal influence, and the influence from the market growth all have positive and significant effects on diffusion, though the external influence is very small. The diffusion of the ADSL subscribers of the NTT group is mainly determined by the internal influence and the growth of the total market. The results also suggest that the diffusion patterns differ between the introduction stage and the growth stage. In the introduction stage, the influence of the total market's

growth on NTT was twice as large as that of the Others. Since the internal influence for each provider group was assumed to be identical and the external influence is very small, the larger value of the parameter r_2 means the rapid growth of the NTT group. In the growth stage, the difference of the influence of the total market's growth on the NTT group and the Others became small. This is probably contributed to the entry of Yahoo!BB as the principal member of the Others.

Using the estimated diffusion parameters, the estimated diffusion patterns are computed separately for the NTT group and the Others. Figures 2 and 3 represent the estimated and the actual diffusion patterns for the NTT group and the Others, respectively. Comparing the diffusion patterns of the estimated with the observed, it is obvious that the proposed brand-level diffusion model fits to the ADSL data in Japan very well.

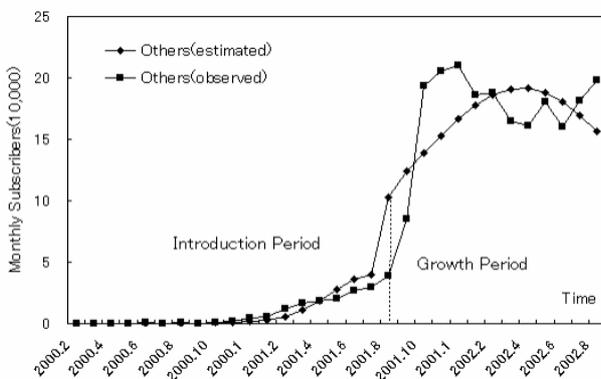


Figure 2 Observed and estimated subscribers (Others)

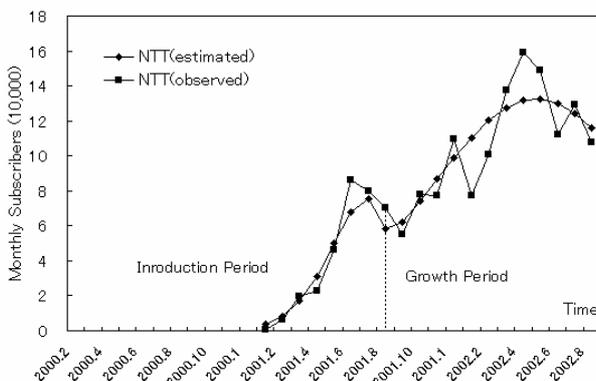


Figure 3 Observed and estimated subscribers (NTT)

5. CONCLUSIONS AND MANAGERIAL IMPLICATIONS

This work would be the first attempt to capture the diffusion dynamics by the three forces – the external influence, the internal influence, and the influence from the growth of the total market. Since the Bass model (Bass, 1969) was introduced into the marketing literature, a number of researches about innovation diffusion used the external coefficient p and the internal coefficient q to describe and explain the difference of innovation diffusion patterns. Empirical results across

different products and markets revealed that the Bass model fitted to the innovation diffusion data very well. Undoubtedly, using parameters p and q , the Bass model can successfully capture the diffusion dynamics when a particular product category is under consideration. However, managers are much more interested in the adoption or growth process of a particular brand and the competitive relationship among brands in a product category. One of the advantages of the Bass model is that it depends on the communication pattern underlying the diffusion process – the external communication through mass media, and the internal communication with the early adopters of the product. When we consider the communication pattern for a particular brand in a product category, not only the mass media and the early adopters of a particular brand, but also the growth of the total market would influence the diffusion of that brand. For some products or services such as the cellular phone and the ADSL service, the primary advantages of adopting these products or services are not so much different depending on the competing providers. Therefore, the influence of the growth of the category on the growth of each brand may be large. In the case of other products such as cars, the market of which is characterized as a completely differentiated market, the success of the total market may only have little impact on the adoption of a particular brand.

The principal implication of this study is the consideration of the category growth effect on the adoption of a particular brand to describe the brand level diffusion process. In the concept of the proposed model, the influence of the category growth on each brand can be different, and captured by the value of the parameter r . The empirical analysis revealed that the category growth effect on each brand was different, and varied with the growth of the market. The later entry may have the advantages of facing a established market with familiarity with the innovation, confidence about the innovation, and preparation for its adoption. In this study, the entry of NTT made up the ADSL market category and several brands grew up in the market. This also happened when Yahoo!BB entered into the market by using low pricing strategy. The success of the NTT and Yahoo!BB in the ADSL service market implied that the timing of entry is very important. NTT entered into the market about one year after the leaders such as Tokyo Metallic Communication and eAccess entered into market, when the ADSL had obtained high recognition. Yahoo!BB entered into the market six months after the NTT, and expanded its market share through large price reduction.

Through assuming that the internal influence parameter of each brand in a product category is identical, the proposed model can be aggregated to the category Bass model. This assumption brings the advantage of the proposed model that it can be solved in a closed-form expression as a function of time t only. Researches about the brand level innovation diffusion have

employed discrete models to describe continuous phenomenon (Mahajan, Sharma, Buzzell 1993; Parker and Gatignon 1994; Givon, Mahajan, Muller 1997, Modis 1997). Although the discrete method make it possible to easily handle the problems and test various assumptions about the effect of competition among brands, it has fatal shortcoming of estimation biases and sometimes it cannot provide the information on the estimation errors. Using the closed-form solution of the proposed model, the estimates and their standard errors can be obtained directly.

REFERENCES

- [1] Bass, F. M., "A New Product Growth Model for Consumer Durables," *Management Science*, Vol. 15, No. 5, pp.215-227, 1969.
- [2] Chatterjee, R., J. Eliashberg, V. A. Rao, "Dynamic Models Incorporating Competition," *New Product Diffusion Models*, edited by Mahajan, Muller, Wind, pp.165-205, 2000.
- [3] Gatignon, H., J. Eliashberg, T.S. Robertson, "Modeling Multinational Diffusion Patterns: An Efficient Methodology," *Marketing Science*, Vol. 8, pp.231-237, 1989.
- [4] Givon, M., V. Mahajan, E. Muller, "Software Pirate: Estimations of Lost Sales and the Impact on Software Diffusion," *Journal of Marketing*, Vol. 59, pp.29-37, 1995.
- [5] Givon, M., V. Mahajan, E. Muller, "Assessing the Relationship Between the User-Based Market Share and Unit Sales-Based Market Share for Pirated Software Brands in Competitive Markets," *Technological Forecasting and Social Change*, Vol. 55, pp.131-144, 1997.
- [6] Krishnan, T. V., F. M. Bass, V. Kumar, "Impact of Late Entrant on the Diffusion of a New Product/Service," *Journal of Marketing Research*, Vol. 37, pp.269-278, 2000.
- [7] Mahajan, V., E. Muller, F. M. Bass, "New-Product Diffusion Models," *Handbooks in Operation Research and Management Science*, Vol. 5, edited by J. Eliashberg and G. Lilien, pp.349-408, 1993.
- [8] Mahajan, V., S. Sharma, "Simple Algebraic Estimation Procedure for Innovation Diffusion Models of New Product Acceptance," *Technological Forecasting and Social Change*, Vol. 30, pp.331-346, 1986.
- [9] Mahajan, V., S. Sharma, R. D. Buzzell, "Assessing the Impact of Competitive Entry on Market Expansion and Incumbent Sales," *Journal of Marketing*, Vol. 57, pp.39-52, 1993.
- [10] Modis, T., "Genetic Re-Engineering of Corporations," *Technological Forecasting and Social Change*, Vol. 56, pp.107-118, 1997.
- [11] Nakajima, N., "A New Diffusion Model in a Competitive Environment and Its Application to an Entry Timing Decision Problem," *Osaka Economic Papers*, Vol.40, pp.402-410, 1990.
- [12] Parker, P., H. Gatignon, "Specifying Competitive Effects in Diffusion Models: An Empirical Analysis," *International Journal of Research in Marketing*, Vol. 11, pp.17-39, 1994.
- [13] Peterson, R., V. Mahajan, "Multi-Product Growth Models," *Research in Marketing*, JAI Press, Inc. Vol. 1, pp.201-231, 1978.
- [14] Rogers, E. M. (1995), *Diffusion of Innovations*, 4th. Edition, Free Press: New York, 1995.

Neural Networks Based Integrated Evaluation Method for the Effectiveness of CRM

Qingliang Meng, Qinghua Kong, Yuqi Han, Jie Chen

School of Economic & Management, Nanjing University of Science and Technology, Nanjing 210094, China
mengzhi007@sina.com

ABSTRACT

Customer relationship management (CRM) has become one of the leading business strategies in the new economy. The effectiveness of CRM can be measured as a satisfaction level achieved by CRM activities. CRM has emerged as a major business strategy for e-business, but little research has been done in evaluating the effectiveness of CRM because of its complexity. In this paper, on the basis of building a CRM evaluation index system, we propose neural networks based integrated evaluation method for the effectiveness of CRM. It can simulate evaluation made by experts and avoid subjective mistakes. The results from the simulation are satisfied.

Keywords: customer relationship management, neural networks, integrated evaluation

1. INTRODUCTION

Transforming enterprises to become customer centric to get more revenue and profit is one of the hottest strategies in business today. This strategy is known as customer relationship management (CRM). As enterprises pursue CRM strategies, they become aware that the costs and benefits of CRM initiatives are significant, and they strive to grasp the financial impacts and economic factors that contribute to their success. For enterprises to achieve ROI (return on investment) from CRM, investments in the application domains and technologies of CRM should contribute tangible business benefits to the enterprise, as well as intangible benefits. The effectiveness of CRM can be measured as a satisfaction level achieved by CRM activities. As CRM has emerged as a major business strategy for e-business, evaluating its effectiveness is very important. However, little research has been conducted to evaluate CRM effectiveness. In the CRM world, increased emphasis is being placed on developing measures that are customer-centric and give managers a better idea of how their CRM policies and programs are working [1].

In this paper, we develop neural networks based integrated evaluation method for the effectiveness of CRM. Because artificial neural network (ANN) acts as a model of real world system or function. The model then stands for the system it represents, typically to predict or to control it. Among the various artificial neural networks presented so far, the multi-layer feed forward neural network is employed in this study. It is found that the ANN modeling is very effective and accurate.

2. CUSTOMER RELATIONSHIP MANAGEMENT

CRM can be defined as managerial efforts to manage business interactions with customers by combining business processes and technologies that seek to understand a company's customers. Companies are becoming increasingly aware of the many potential

benefits provided by CRM. Some potential benefits of CRM are as follows:

- (1) Increased customer retention and loyalty
- (2) Higher customer profitability
- (3) Creation value for the customer
- (4) Customization of products and services
- (5) Lower process, higher quality products and services

3. A FRAMEWORK FOR EVALUATING THE EFFECTIVE NESS OF CRM

3.1 The CRM process model

Because it is difficult to evaluate tangible returns on the resources expanded to plan, develop, implement, and operate CRM, we have to measure the intangible nature of benefits such as customer loyalty, service quality, value enhancement, effectiveness of processes, innovation of operation, service improvement, competitiveness, trust, and efficiency. To evaluate the effectiveness of CRM, we need to build a evaluation framework, it must be a multi index system and it must can evaluate the whole CRM process. In figure1 we give the model of CRM process.

From figure1 we can see. First, a company accumulates a huge amount of customer information and creates a customer profile. Next, the company discovers large and hidden customer characteristics through data mining tools and techniques. As CRM has tremendous potential for collecting and storing customer preferences, CRM can make it possible to create new products and customize existing products in innovative ways. Second, the company integrates all relevant information on each customer across the enterprise in order to facilitate more effective planning, marketing and services. Identified customer knowledge helps to find customer needs when a company interacts with customer. As customer requirements and expectations are met, loyal customers are created. Customer value can be added by product and service customization, additional information provision,

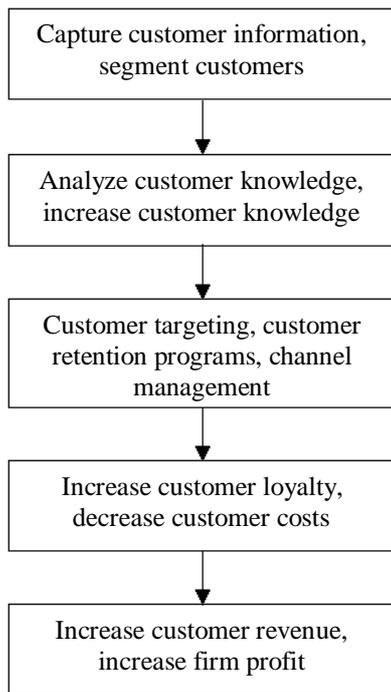


Figure 1. The model of CRM process

and quality enhancement. Understanding and collecting customer needs are critical to conducting these value-adding activities. Third, satisfactory long-term customer relationships can be established, as channels aid employees to serve customers effectively and profitably. Finally, the company increases revenues and profits. Typically, business value can be added by improving brand image, establishing customer relationships, enhancing public relations, and generating sales leads [2].

3.2 The evaluation index system for effectiveness of CRM

The model of CRM process can give a clue as to what perspective are important to evaluate the effectiveness of CRM. That is customer knowledge, customer interaction, customer value, and customer satisfaction. So an evaluation index system can be shown in figure 2.

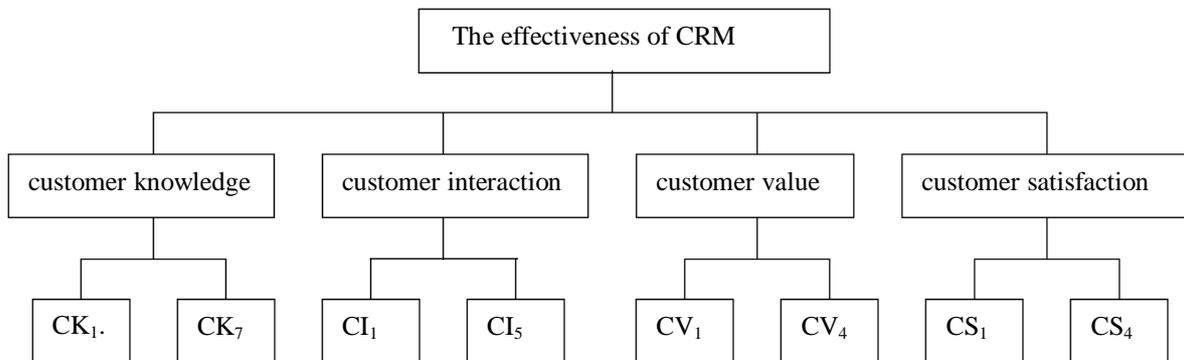


Figure 2. The evaluation index system for effectiveness of CRM

Where, CK₁: collecting appropriate customer information;
 CK₂: analyzing customer data;
 CK₃: acquiring new customers;
 CK₄: understanding customer needs;
 CK₅: improving skills of employee;
 CK₆: improve CRM techniques;
 CK₇: secure service;
 CI₁: appropriate response to customer request;
 CI₂: integration of business process;
 CI₃: improving channel management;
 CI₄: maximizing the effectiveness and efficiency of enterprise operations;
 CI₅: customizing products and services;
 CV₁: improving customer retention;
 CV₂: profits increase;
 CV₃: improving customer service and support;
 CV₄: building attractive virtual community;
 CS₁: improving service quality;
 CS₂: establishing relationship with customers;
 CS₃: responsiveness;
 CS₄: reliability.

The customer knowledge (CK) perspective represents the status of the customer segment and customer data management. Customer knowledge focuses on technology learning, understanding customer needs, and customer profiles, which influence ways of interacting with customers. The varied tastes and preferences of customers make it difficult to classify customers into a large homogenous population to develop marketing strategies. A customer wants to be served according to his or her individual and unique needs. To analyze customer needs, CRM uses appropriate data mining tools and data warehousing techniques. Technology learning is also important towards understanding customers. It is required, therefore, to assess employee skill to use customer information effectively. Security is another basic and critical prerequisite when dealing with customer information [2]. Security, in particular, has been a serious issue concerning online purchases and an impediment to the acceptance of the e-channel. Many customers are concerned about the amount of personal information that is contained in databases and how it is being used. Customers perceive safety of transactions and seller empathy as important.

The customer interaction (CI) perspective represents operational excellence and channel management of customer services and management processes. To manage various communication channels efficiently, managers make an effort to monitor the business processes. Winer suggested a relational program that includes customer service, loyalty programs, customization, rewards programs, and community building [1]. To analyze customer interaction, some important measures need to be considered, such as the number of marketing campaigns, total cost for promotion, frequency of contents updates, payment, response channels, and so on. Communication channels not only include classic communication channels such as letters, fax, and telephone but also emerging new channels such as call centers, service centers, Web sites, and virtual internet communities. It is vital to manage various channels efficiently and immediately.

The customer value (CV) perspective represents the benefits gained from customers, such as lifetime value. If the customer is satisfied with the service of a company, there will be no problem concerning customer deviation. To maintain this relationship, the CV perspective continues to find ways to build customer commitment and loyalty. In addition, customers are identified by their value and are treated accordingly. Customer value describes tangible and intangible benefits gained from CRM activities, which help to arrange the relationship with the customer successfully. In order to determine the customer value, we need to analyze such information as customer retention, profits increase, customer service and support, virtual community. Current customer profitability should be calculated, establishing a baseline and comparing new calculations to that baseline periodically. Calculating customer value potential and using it as a guideline will be profitable in the future [3].

The customer satisfaction (CS) perspective represents the level of satisfaction achieved by products or services. Customer satisfaction is the feeling that a product or service meets customer expectations and determines whether the buyer will become a permanent customer or not. The assessment of customer satisfaction is one of the most important stages of the implemented model. Customer satisfaction is difficult to measure because it is hard to quantify the satisfaction level [4]. Customer satisfaction represents a modern approach for quality in enterprises and organizations, and serves the development of a truly customer-focused management and culture. Measuring customer satisfaction offers an immediate, meaningful, and objective feedback about customer preferences and expectations. Among the four perspectives, the customer satisfaction perspective is the most important because customer satisfaction is directly linked to a company's profits. Customer satisfaction can be measured from the four dimensions: service quality, the relationship between enterprises and customers, responsiveness, reliability. Responsiveness is the willingness to help customers and provide prompt

service [5]. Reliability is the ability to perform the promised service dependably and accurately .

4. NEURAL NETWORKS BASED INTEGRATED EVALUATION METHOD

Artificial neural network algorithm has been used successfully in many applications, because it acts as a model of real world system or function. The model then stands for the system it represents, typically to predict or to control it. ANN can model a function even if the equation describing it is unknown the only prerequisite is representative sample of the function behavior and that is from the experimental data and not from a theoretical understanding. Figure 3 shows the schematic diagram of a multi-layer feed forward network used in this paper. The neurons in the network can be divided into three layers: input layer, output layer and hidden layers.

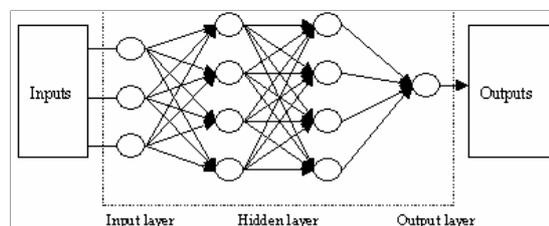


Figure 3. The structure of a multi-layer neural network

Each neuron of the output layer receives a signal from inputs through hidden layer neurons along connections with modifiable weights. The neural network can identify input pattern vectors, once the connection weights are adjusted by means of the learning process. The back propagation learning algorithm which is a generalization of Windrow-Hoff error correction rule is the most popular method in training the ANN and is employed in this work[6]. This learning algorithm is not presented here.

Scaling of the input-output data has a significant influence on the convergence property and also on the accuracy of the learning process. It is obvious that the range of the output of the network must be within $[0, 1]$. The input variables should be kept small in order to avoid saturation effect caused by the sigmoid function. Thus, the input-output data must be normalized before the initiation of training of the neural network. After the normalization, the input variables can then be easily made to fall in the range $[-1, 1]$. Further, the range of the normalized output vector component is made to fall within $[0, 1]$.

There are 20 indexes in the evaluation index system. But each index should have a score made by experts. The scores can be divided into five scales: 1, 0.7, 0.5, 0.3, and 0.1. Echoes best, better, medium, worse and worst respectively. In this example, we take the scores made by experts as the inputs ($ck_1, ck_2 \dots cs_4$) of the neural network model. The expected output is only one neuron,

Table 1. The evaluation data made by experts

order	CK1	CK2	CK3	CK4	CK5	CK6	CK7	CI1	CI2	CI3	CI4	CI5	CV1	CV2	CV3	CV4	CS1	CS2	CS3	CS4	score
1	0.7	1	1	0.7	0.7	1	0.7	0.7	1	1	0.7	1	0.7	0.7	0.7	0.7	1	1	1	1	0.861
2	0.5	0.7	0.5	0.7	0.7	0.5	0.5	0.7	0.5	0.7	0.7	0.5	0.7	0.5	0.7	0.7	0.7	0.7	0.7	0.5	0.604
3	0.5	0.5	0.3	0.5	0.3	0.3	0.3	0.5	0.3	0.5	0.3	0.3	0.3	0.5	0.3	0.3	0.3	0.3	0.3	0.1	0.340
4	0.5	0.5	0.5	0.5	0.7	0.3	0.3	0.5	0.5	0.7	0.7	0.5	0.3	0.5	0.7	0.7	0.7	0.7	0.3	0.5	0.488
5	0.5	0.5	0.5	0.7	0.7	0.51	1	0.7	1	0.7	0.7	1	1	0.7	0.7	0.7	1	0.7	0.7	0.1	0.713
6	1	0.7	0.7	1	1	1	1	0.7	1	1	0.7	1	1	0.7	1	0.7	1	0.3	0.7	1	0.931
7	1	0.7	1	1	1	1	1	0.7	1	1	0.7	1	1	0.7	0.7	0.7	1	1	0.7	0.1	0.810
8	0.7	0.7	1	0.7	0.7	0.5	0.7	0.5	0.7	0.7	0.5	0.5	0.7	0.7	0.7	0.7	0.7	0.7	1	0.5	0.641
9	0.7	0.7	0.5	0.7	0.7	0.5	0.7	0.5	0.7	0.7	0.7	0.5	0.7	0.7	0.5	0.7	0.7	0.7	0.7	0.5	0.683
10	0.7	0.7	1	1	0.7	0.5	1	0.7	1	1	0.7	1	1	0.7	1	1	0.7	0.7	1	0.5	0.827
11	0.7	0.5	1	0.7	0.7	0.5	0.7	0.7	0.5	0.7	0.7	0.5	0.7	0.7	0.7	0.5	0.7	0.7	1	0.5	0.647
12	0.7	0.7	0.5	0.7	0.7	0.5	1	0.7	1	0.7	0.7	1	1	0.7	0.7	0.7	1	0.7	0.7	0.1	0.727
13	0.7	0.7	1	0.5	0.3	0.5	0.3	0.5	0.5	0.5	0.7	0.5	0.3	0.5	0.7	0.5	0.7	0.3	0.3	0.5	0.460
14	0.7	1	0.3	1	0.7	1	1	0.7	1	1	0.7	1	0.7	0.7	0.7	0.7	0.7	0.7	1	0.5	0.817
15	0.7	0.7	0.5	1	0.7	1	1	0.7	1	1	0.7	1	1	0.7	0.7	0.7	0.7	0.7	0.7	0.5	0.766

and it must be within [0,1], and it represents the result of the evaluation. The result scales can be divided into four levels: if the value is between 0.7 and 1, we treat it as good, if the value is between 0.5 and 0.7, we treat it as medium, if the value is between 0.3 and 0.5, we treat it as poor, if the value is below 0.3, we treat it as very poor. In next section, we will develop a simulation from an example.

5. SIMULATION AND RESULTS

In this paper, we will use the data acquired from 15 firms, it is shown in table 1. The scores of each index were made by experts. The scores of last column are the results achieved by experts using multicriteria integrated evaluation method. We can use neural networks based integrated evaluation method to simulate. In this example, we choose the former 12 samples as training signals, and the left 3 samples as test signals. Then we use Matlab as the tool, leading rate is 0.0001, the number of hidden neurons is 7, the training times is 1500, weight adjusted parameters is 0.5. The results can be obtained in table 2.

Table 2. The learning results

Order	Results of training	Expected outputs	Errors / %
1	0.8583	0.861	0.270
2	0.6008	0.604	0.320
3	0.3394	0.340	0.060
4	0.4726	0.488	1.540
5	0.7078	0.713	0.520
6	0.9274	0.931	0.360
7	0.8057	0.810	0.430
8	0.6420	0.641	-0.100
9	0.6739	0.683	0.910
10	0.8253	0.827	0.170
11	0.6413	0.647	0.570
12	0.7257	0.727	0.130

We can see it is very similar to the expected outputs. The test results are shown in table 3. The error variation curve is shown in figure 4. The results of evaluation are consistent with the scores made by experts. Thus, the simulation is satisfied.

Table 3. The test results and Categorization

order	1	2	3
Test results	0.4508	0.8366	0.6615
Results from experts	0.4600	0.8170	0.6300
Errors / %	1.68	2.37	4.74
Categorized by simulation	Poor	Good	Medium
Categorized by experts	Poor	Good	Medium

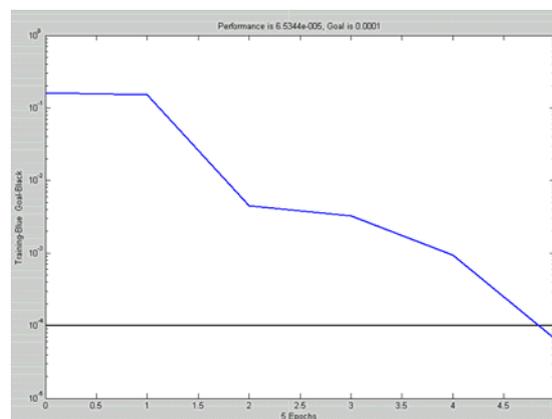


Figure 4. The error variation curve

6. CONCLUSIONS

Customer relationship management is one of the leading competitive business strategies in the new economy. It is very important to evaluate the effectiveness of CRM, because it is beneficial for enterprises to carry out the CRM activities. In this paper, on the basis of building a model of CRM process, we set up an evaluation index system for the effectiveness of CRM, then neural networks based integrated evaluation method are proposed, and the results of the simulation are satisfied.

That is to say, the evaluation method is a powerful assessment tool to evaluate the effectiveness of CRM. We expect that this related work of the method should be further researched so that we can evaluate other fields.

REFERENCES

- [1] Winer, R.S. "A framework for customer relationship management." *California Management Review*, Vol. 43, pp 89–105, 2001.
- [2] Jonghyeok Kim, Euiho Suh. "A model for evaluating the effectiveness of CRM using the balanced scorecard", *Journal of interactive marketing*, Vol 39, No. 2, pp 5-19, 2003.
- [3] Mulhern, F.J. "Customer profitability analysis: measurement, concentration, and research directions." *Journal of Interactive Marketing*, 13 (Winter), pp 25–40, 1999.
- [4] Mihelis, G., Grigoroudis, E., Siskos, Y., Politis, Y. "Customer satisfaction measurement in the private bank sector." *European Journal of Operational Research*, pp347–360, 2001.
- [5] Parasuraman, A. "Reflections on gaining competitive advantage through customer value." *Journal of the Academy of Marketing Science*, 25, pp154–161, 1997.
- [6] Ahmad S. Ahmad, p. s Ghosh, Syed Abdul Kader Aljunid, Halil Hussain, "Artificial neural network for contamination severity assessment of high voltage insulators under various meteorological conditions" *IEEE transactions on electrical insulation*, Vol 24, No. 1, pp 6-11, February, 2003.

Successful Customer Relationship Management: Antecedents and Imperatives

Frederick L. Capossela

International Business & Marketing Department, California State Polytechnic University, Pomona
94-219 3801 West Temple Avenue, Pomona, California 91768 USA
flcapossela@csupomona.edu

ABSTRACT

This paper argues that there are four key conceptual techniques used in Customer Relationship Management (CRM) which should be considered Imperatives for contemporary 21st century e-Business today. Customer Relationship Management itself may be defined as an attempt to find the most efficient way to focus the Marketing effort in order to generate revenue. The antecedents of successful Customer Relationship Management techniques may be found in 20th Century and even 19th Century traditional Direct Marketing constructs. Years of data from this sub-discipline of Marketing, data analysis, and testing confirm the efficacy of these key techniques. Customer Relationship Management concepts which are examined as Imperatives for 21st century e-Business are: 1) Customer Lifetime Value (LTV); 2) Customer Acquisition Cost (CAC); 3) Split-Run (A/B) Testing; and 4) Recency, Frequency, and Monetary Value (RFM) Analysis.

Keywords: e-business, Customer Relationship Management, CRM, Lifetime Value, LTV, CAC, A/B Split Testing, RFM analysis

1. INTRODUCTION

This paper explores the Nineteenth and Twentieth Century antecedents of modern Customer Relationship Management in an attempt to see how they can contribute to greater 21st century e-business profitability.

From a theoretical point of view, Customer Relationship Management (CRM) can be seen as a more efficient focus of the effort to generate revenue—the task of Marketing. The reason that Customer Relationship Management has been gaining prominence in today's economy is that it has demonstrated success at generating revenue more efficiently. [1]

The current practice of Customer Relations Management today confirms the prescience of Peter Drucker when he wrote in 1973:

To know what a business is we have to start with its purpose. Its purpose must lie outside of the business itself. In fact, it must lie in society since business enterprise is an organ of society. There is only one valid definition of business purpose: to create a customer. [2]

The foundation of CRM success today is the concept of Customer Lifetime Value (LTV). Lifetime Value may be thought of as an elaborate extension of Pareto's principle.[3] (Pareto's maxim that eighty percent of an organization's revenues will be generated by twenty percent of its customers has been well accepted for decades.)

2. ANTECEDENTS

However, Lifetime Value is but one of a series of concepts whose success has been proven and refined by a sub-discipline of marketing called Direct Marketing. In 1872, American merchant, Montgomery Ward, and later, Richard Sears sent their catalogs to a predominantly rural population (9 out of every 10 Americans lived on farms) who were unable to visit a retail store. [4] Since that time the field of Direct Marketing has developed a number of operational concepts that have grown in sophistication, power, and profitability through a range of media including catalogues, direct mail, telemarketing, and direct response advertising.

The Direct Marketing concepts which have the most to offer to today's e-commerce practitioners are the ones identified as imperatives in this paper.

The success and wisdom of Direct Marketing has been earned over the better part of a century. The principle reason is that Direct Marketing also known as Database Marketing, enjoys such success is its 100% accountability. In contrast to mass media promotion, Direct Marketing identifies exactly whom it wishes to target (individuals by actual name). This allows collection of data on exactly who responded yielding 100% accountability across a range of metrics. For example, if 100 catalogues are mailed to 100 names, the direct marketer knows the names of the 10 who responded and exactly what they bought and exactly how much each sale cost.

In addition to Lifetime Value (LTV) other operational concepts which Direct Marketing has developed and implemented over the years include Customer Acquisition Cost (CAC); Split-Run Testing (A/B); and

Recency, Frequency and Monetary Value Analysis (RFM). Each of these constructs can be adopted for use by e-business. Since all the techniques can be implemented and analyzed quantitatively, they can easily be applied across language and culture making them ideal for global application.

Because these antecedents have been so reliable and successful in the past for Direct Marketing, their promise for modern e-Business is obvious. Once fully appreciated, these Direct Marketing antecedents may be regarded as 21st century Customer Relationship Management (CRM) Imperatives.

3. IMPERATIVES

The first of four antecedents of the practice of Customer Relationship Management (CRM) which should be regarded as imperatives is the concept of Lifetime Value. This term posits that there is a lifecycle to the relationship that customers have with an organization or business. During the period of the relationship between the beginning and the end, the company has an opportunity to lengthen or shorten the relationship based on how it “manages” the relationship. “Manages” is an unfortunate choice of words since it indicates an almost adversarial nature of the relationship, or at the least, an unequal one, between the manager and those managed. Nevertheless, Customer Relationship Management is the term in accepted use today.

3.1 LifeTime Value (LTV)

Enlightened businesses realize that every time a customer has an interaction with the company (or “touch” in CRM language), the relationship can be either be deepened or weakened. If the interaction is a positive “touch,” there is an opportunity for cross-selling and increased profit. An additional benefit of deepening trust and frequency is loyalty. Loyalty extends the term of the relationship so that a customer who has been with a company five years continues for six, eight, or ten years. This extension not only increases the net present value (NPV) of the projected revenue stream, but does so at lower cost. This in turn, increases margins and leads to greater profitability. [5]

Any calculation of Life Time Value (LTV) must take three factors into consideration:

1. Margin
2. Retention Rate
3. Discount Rate

First, Margin is the annual revenue that customers generate minus the company’s operating expenses in serving them. Second, the Retention Rate is the percentage of first-year customers who can be expected to continue into the second, third or n year. And third, the Discount Rate is the current cost of capital. [6]

LTV tracks the buying behavior of a group of customers and projects it into future years. In order to compute LTV, we need to know the following: [7]

- number of customers acquired in first year of business
- cost of promotional materials needed to acquire them
- cost of promotional materials after their first purchase
- retention rate (first year customers who continue into second year)
- average order size
- number of orders per year
- cost of goods sold
- average rate of returns, exchange, and damages
- market rate of interest

The Imperative is that once Lifetime Value has been computed for each customer, an analysis can be made to identify those *customers who represent greater value to a company’s future* and those who represent less. Decisions can then be made as to which customers are the company’s best prospects, which are marginal and which are unprofitable to the company. This analysis helps decision-makers formulate promotional strategies which can most efficiently grow a company’s profitability. Managers can now make informed decisions as to whether to concentrate on customer retention, acquisition, or win-back. [8]

3.2 Customer Acquisition Cost (CAC)

The next antecedent that is integral to implementation of Customer Life Time Value (LTV) is the accurate determination of Customer Acquisition Cost (CAC).

Reichheld claims that “On average, U.S. corporations tend to lose half their customers in five years, half their employees in four, and half their investors in less than one.”[9] The implication of these facts, underscores the continuing requirement for companies to replenish their customers. The cost of new customers is seldom free. However, not all customers cost the same to acquire. Nor do acquisition programs always yield the same quality customers. Moreover, customer acquisition costs can vary wildly.

Calculating Customer Acquisition Cost is a technique that can assign a dollar cost to acquiring or re-acquiring each customer, old or new. A Customer Acquisition Cost ratio is used to help determine whether to retain an existing customer or to acquire a new one.

In hindsight, it can be seen that among the many factors responsible for the dot.com bust was lack of appreciation of CAC. At the height of the dot.com boom, to give just one example, the company, e-Toys, was reportedly spending more than \$700 to acquire each new customer. Unfortunately, each new customer’s LTV was only \$87. The inevitable result was that for

every new customer acquired, e-Toys would absorb a loss of over \$600.

To truly appreciate the value of keeping track of CAC, there is one informative general Direct Marketing principle to apply. The rule-of-thumb used by direct marketers is that it costs between 5 to 10 times more money to acquire a new customer than it costs to retain a current customer. [10]

Of particular importance is the fact that the Customer Acquisition Cost (CAC) for pure-play Internet retailers is higher than their multi-channel counterparts. Internet retailers are defined as those with physical stores or mail-order catalogs in addition to their web sites. [11]. CAC is defined as total sales and marketing expenses divided by new customers accounts. Author Morrison predicts that average Customer Acquisition Cost will rise to \$150 for web-based retailers vs. \$95 for multi-channel retailers in 2004.

One of the most difficult problems associated with computing true CAC is that it becomes difficult to distinguish between which sales and marketing programs stimulate repeat sales to existing customers and which result in new customer acquisition.

The Acroterion Company, estimates that CAC for pure-play retailers is \$82 per customer while a multi-channel retailer's is only \$31 per customer according to a survey of 221 retailers conducted by Shop.org. [12]

Compared to a multi-channel retailer, the pure-play retailer has a much more difficult task in identifying and reaching potential customers. While direct marketers can choose from more than 1,000,000 databases of names and addresses of groups segmented by demographic, geographic, and psychographic variables, the pure-play retailer has a much more difficult job. The fact is there are so many fewer lists available containing valid e-mail addresses. In this day of rampant "spam" and extraordinary privacy concerns, direct solicitation of potential customers via email no longer appears viable. Consequently, pure-play retailers must resort to more expensive methods like on-line advertising and search-engine listings.

3.3 Spit-Run (A/B) Testing

The next antecedent is key to Direct Marketing's success, the concept of Testing. Direct Marketers test every variable they can in an effort to achieve optimization in their choice of merchandise, their value proposition, and the elements of their promotional materials.

E-Commerce practitioners have much to learn in this area. First, there is the scope of what to test – everything that is important. And, second, there is the relentless

dedication to testing, the never-ending emphasis on optimization.

One of the most frequent statistics mentioned about e-commerce shopping is that 7 out of every 10 shopping attempts on line are abandoned. Clearly, this is unacceptable. It is an area ripe for improvement. It is here that ingenious testing can make a significant contribution.

The term that direct marketers use to describe their testing technique is Split-Run (A/B) Testing. On the Internet, it can be used to test such things as product features, web page design, and navigation, value. [13]

In its simplest form, the e-commerce retailer selects one--and only one--variable at a time to be tested, such as the number of steps necessary for a successful shopping cart check-out. Assume that seven is the current number of steps, called the "control." A new 6-step sequence is now constructed. Then, a method is developed so that every other visitor to the shopping site will be offered the 7-step checkout and every other visitor will get the new 6-step sequence.

By tracking how many shoppers have used the 6-step vs. how many have used the 7-step sequence, we can determine whether reducing the number of steps has any significant effect on the outcome. If the new 6-step sequence increased the number of users, then it "wins" the test and becomes the new "control" and the 7-step sequence is retired.

The imperative here is that e-commerce retailers must constantly search for variables to test and be relentless in their commitment to continual testing. Moreover, they must commit to a continual refinement of all aspects of marketing program in keeping with those test results.

The Internet offers the extraordinary opportunity for dynamic, real-time testing that traditional direct marketers can only dream about. For example, in a 24-hour period, an e-commerce retailer can vary the size of a picture of an item offered on a web page whereas a traditional direct marketer using a picture on a page in a printed catalog would only be able to compare results from the control, the current size, Size A, and the new size picture, size B. And then the retailer would have to wait to reprint the catalog in order to test again.

3.4 Recency, Frequency, Monetary Value Analysis

The final antecedent is called RFM analysis which means analysis of the Recency, Frequency, and Monetary Value of customer purchases.

RFM analysis is a refinement of Pareto's principle, commonly known as the 80/20 principle, which

translates into marketing operations as the axiom that 80% of an organizations revenues will be generated by 20% of its customers.

Recency, Frequency, and Monetary Value (RFM) Analysis predicts customer behavior. [14] Attention to RFM delivers empirical proof of a company's best customers. Moreover, it predicts exactly which promotional solicitations will be most successful by market segment. Each record of an existing database is coded to identify which customers bought most recently, which customers bought most frequently, and which customers spent the most money on the company's goods and services.

At first, RFM analysis appears counterintuitive, because it is based on the principle that customers who bought most recently will be the most likely to buy again. However, the proof of this has been demonstrated empirically time and again by Direct Marketers. Likewise, the principles of frequency of purchase and dollar amount spent have also been demonstrated to be empirically sound. Marketers code all customers into RFM cells and examine the response rates of customers in each cell when exposed to the same promotion. Generally, Recency is weighted at 50%, Frequency at 30%, and Monetary Value at 20%. [15] Chi square analysis is used to analyze the data and produce results.

3.4.1 Migration Analysis

A refinement of RFM analysis is called Migration Analysis. Results of migration analysis indicate patterns of customer migration, which allow the user to predict which cells customers will migrate to in the future. This, in turn, will permit the company to focus promotional efforts against the most likely and receptive segment cells. Conversely, those customers (cells) that will be most likely to migrate into lower ranking cells can similarly be predicted. [16]

The imperative here is that RFM analysis is based on data that is readily available to e-Businesses. Knowing the usefulness of such data is ample incentive to gather it consistently. Once RFM analysis becomes a central part of the Customer Relationship Management process, it can become a powerful tool which e-Businesses can implement to increase their profitability.

4. CONCLUSION

Finally, as modern, traditional marketing emerged during the second half of the 20th century, its emphasis was on mass distribution and mass media promotion. This type of marketing largely ignored the valuable lessons of one-to-one marketing that its earlier sub-discipline, Direct Marketing, had learned and refined.

Today, there has been an explosion of one-to-one marketing opportunities thanks to the World Wide Web, making the tried and tested lessons of Direct Marketing imperative for any e-business to study. The reason is simply that the business model closest to successful 21st century e-business is not traditional 20th century "modern" marketing, but venerable Direct Marketing. It is this body of experience that holds the greatest promise for creating a sustainable competitive advantage and greater profitability for e-Business.

5. REFERENCES

- [1] Rust, R., Zeithami, Lemon, *Driving Customer Equity: How Customer Lifetime Value is Reshaping Corporate Strategy*, New York: Simon & Schuster, 2000.
- [2] Drucker, P., *Management Tasks, Responsibility, Procedures*, New York, Harper Business, p. 61, 1973.
- [3] Pareto, V., *The Mind and Society*, New York: Harcourt, Brace & Co, 1935.
- [4] Crossen, C., "Mailed 'Wish Books' Let Americans Shop Beyond their Needs," p. B1, *Wall Street Journal*, December 17, 2003.
- [5] Peppers, D., Rogers, M., *Managing Customer Relationships*, New York: John Wiley, P. 5, 2004
- [6] "The Real Value of Customer Loyalty," *MITSloan Management Review*, Vol.43, No. 2, pp14-17, Winter 2002.
- [7] Hughes, A., "The Value of Lifetime Value," *Catalog Age*, March 2004.
- [8] Gupta, S., Lehmann, D., Stuart, J., "Valuing Customers," *Marketing Science Institute Report*, 01-119, 2001
- [9] Reichheld, F., "Learning from Customer Defections," *Harvard Business Review*, pp.87-88, March/April 1996.
- [10] Laszlo, J., "Sacrificing Retention on the Altar of Acquisition," *jupitaresearch.com*, September 24, 2002.
- [11] Morrison, E., "The Cost of Finding Customers," *iQ Magazine*, November/December 2001.
- [12] "Knowledge Base, Customer Acquisition Cost," *Acroterion.com*, August 2004.
- [13] Rogers, M., "Deeper Profiles Lead New Database Efforts," *Inside 1to1 Magazine*, July 2004.
- [14] Schell, E., "Date Tech: The Case for Database Marketing," *Catalog Age*, February 2003.
- [15] Stone, R., & Jacobs, R., *Successful Direct Marketing Methods*, 7th Ed., New York: McGraw-Hill, 2001, p. 10.
- [16] Sellers, J., Hughes, A., "RFM Migration Analysis: A New Approach to a Proven Technique," *Database Marketing Institute*, August 2004.
- [17] Peppers, D., Rogers, M., *Managing Customer Relationships*, New York: John Wiley, p. 5, 2004.

Supporting Small Businesses' Penetration of New International Markets - Practitioner's Approach.

Jaroslav Jandos

Faculty of Informatics and Statistics, University of Economics, Prague 130 67, Czech Republic
jandos@vse.cz

ABSTRACT

Small businesses typically have limitations in areas of modern marketing and the use of ICT and I-commerce (Internet-commerce) in particular for penetration of new (mainly international) markets. An approach that can be used to address these limitations – i.e. method and implementation concepts based on the use of external consulting team (including University students) - is presented in this paper. Our approach is practically oriented and has attractiveness for small businesses in that it typically can provide practical results in relatively short time (several months) with reasonable costs. Practical usefulness of this approach, which is also an example of University–Industry partnership, is illustrated by implementation of method in a small Czech business.

Keywords: I-marketing, small business, SME, 1:1 marketing, CSF

1. INTRODUCTION

Small businesses often aim at penetration of new, often international, markets. For this purpose not only traditional means and ways (methods) but also means of ICT (Information and Communication Technologies) and corresponding methods are used. Small businesses often don't have own employees with required qualification – especially in areas of marketing, ICT (including Internet) etc. This is the main reason why above mentioned activities of small businesses are often executed with assistance of external consulting team. These services are often expensive and project often lasts long. For these reasons it seems reasonable to support these small businesses activities by consulting team, which includes University students. They often have reasonable qualification and can join project for fraction of costs of professional consultants.

“Small business” includes not only SME (Small and Medium Enterprise -classification of which can be different in different countries) but also other enterprises, which can find presented approach as valuable.

Although our approach is not limited to small businesses it has a special attractiveness for them in that it typically can provide practical results in relatively short time (several months) with reasonable costs.

Paper is oriented towards support of small business's expansion to new (mainly international) markets using ICT, namely Internet and Internet marketing and Internet commerce.

This support was executed as project with specified duration, resources and goals. Project was executed in small business using consulting team (including University students) with following main goals

- development of method for finding new (mainly international) markets for small business's penetration, resulting in new contracts
- demonstrate to small businesses the advantage if ICT's (mainly Internet-based) use in internal and external commercial activities
- training of small business's employees in novel approaches and use of ICT in such a way, that they can use these means further even after conclusion of the project

Paper topics include namely

- development of simple practical method for support of small business's penetration of new markets and it's practical execution as project
- University - Industry partnership, by inclusion of University students into consulting team
- knowledge transfer between members of consulting team and small business's employees during the Project

2. ROLE OF CONSULTING TEAM

The method (see below) is based on use of consulting team consisting typically of experienced professional (acting as project manager) and 3 to 4 university students

- consulting team is group of external experts, who execute those activities for which small business does not have own qualified employees. These are especially areas of marketing, e-commerce, use of ICT in business, project management.
- from the point of view of perspective customers/partners consulting team works fully “in the background” i.e. all external commercial activities are carried out by small business
- consulting team members can be permanently located in different geographical localities (in one country, even in different countries). They collaborate on reaching the objectives and deliverables of methodology for small business. Based on project plan they participate in

specified activities (submission of deliverables to small business's approval etc.) with small business management. For these reasons the use of convenient methods and technologies for collaboration is important.

Consulting team completes the following tasks

- manages individual steps of method, which are implemented as processes, proposed and managed by consultant. At the same time, the current small business processes are executed, i.e. project processes are mapped "on top" of current processes. After project termination small business can decide to adapt some of these project processes as his new processes, which are now executed by small business's own employees. Small business appoints own employees (typically with limited qualification) for execution of particular project steps
- proposes structure and the way of project management, project goal, expected project resources. After approval by small business this is the basis of the project
- proposes partial methods and approaches used in individual steps of project, ICT used in project as well as methods of communication and collaboration in project
- for particular tools/methods used in project (e.g. Internet market research using search engines, collaboration technology) provides training of small business employees (at least of those who take part in project) so that after project termination they can use these tools by their own.
- provides (with small business's assistance) actual execution of those activities for which small business does not have own qualified employees

3. METHOD – BASICS

For formulation of method for supporting small business expansion to new international markets mainly the following factors are taken into account

- method is based on known concepts of marketing, I-marketing, ([1], [2], [4], [8]) e-commerce, I-commerce ([5], [6], [9]). These concepts are applied in such a way that concrete benefits for small business are reached within reasonable time (several months) and with reasonable costs. Approach to method formulation is not governed by full-width theoretical approach but more by an attempt to formulate relatively straightforward method, which could be implemented relatively quickly under conditions, which are typical for most small businesses. The core of the method is the discipline of Internet marketing.
- it is assumed that small business typically does not have own employees with required qualification in areas like ICT, marketing, I-commerce. This is (on the basis of our practical experience) particularly true for smaller businesses. Therefore the implementation of method with the use of external consulting team is typical. In case small business has own qualified employees it can implement method by its own staff.

- Implementation of method is considered to be a project, with all required attributes. Project is typically managed by head of consulting team. Project team includes consulting team members as well as small business's employees

3.1 Marketing – Basic Approach

Typical marketing activities are executed as steps of the method. However specific approaches have been selected for their implementation. These are namely the following

- Internet advertising (banners, search engine marketing etc.) was not used at all. Only active approach to market analysis was adopted. ICT based methods were primarily used – heavy use of Internet for analysis of company web sites, B2B info portals, e-marketplaces (as source of info) as well as search engine use – supported by traditional methods – analysis of printed materials acquired by project team, phone contacts etc. Info on particular potential business partner ("lead") was made more accurate until the responsible person was found, who was then contacted
- For ethical reasons the responsible person at potential partner was typically first contacted by short phone call in order to find out whether he/she agrees with product info being sent to him/her (i.e. "permission-based promotion")
- Marketing info including product info was delivered to potential partners typically by e-mail, company web site (only very basic info – this web site was operational before start of the project) or by 1:1 presentation in separate rooms of Quickplace (for potential partners with the greatest "value" – see below) which was (in step 6 of the method) further made more accurate so that it reacts to the exact requirements of particular partner

In the course of marketing activities several strategic decisions has to be made:

- Mass marketing versus direct marketing
Approach of direct marketing was adopted in which only the selected potential partners (included in marketing database – see below) were contacted
- OEM versus brand marketing
Depending on the characteristics of target market as well as on type of potential partner (i.e. wholesale – distributor or integrator) combination of OEM and brand marketing and sales was used on all markets, with OEM having greater success.

Marketing database

Specific marketing database tool was designed and developed as part of the method [3]. It includes up-to-date and relevant marketing meta-information on analyzed markets. This information includes among others – information content, info available on-line or off-line, business subject concerned (customer, competitor, distributor, integrator,...), market segment (i.e. small business's product concerned), region and country, format of primary info (catalogue, on-line

catalogue, corporate web-site,...), importance, contributor (into database).

In the project it's content is updated continuously by members of project team. It is used mainly in steps 5 and 6 of the method by all participants of the project. It was developed in Linux/Apache/PHP/MySQL environment as Web application and made accessible by URL from Quickplace.

3.2 Collaboration Technology

Innovative use of collaboration technology stems from the fact that it is used not only for collaboration but also (and to significant degree) for marketing purposes (see below)

Geographical distribution of project participants (small business, members of consulting team, potential customers) has led to use of collaboration technology. This distribution as well as requirement of asynchronous ("anytime") communication have resulted in the requirement for technology from "different time – different place" quadrant of collaboration technology classification [7]. Lotus Quickplace – Team Workplace, which belongs to category of "web-based team/project rooms" was selected.

Quickplace, implemented in client/serve architecture, requires only Web browser as client, has simple user interface, which makes it very easy to use even by non-specialists. It is based on specification of different "rooms", which can be used by different (groups of) users. Browser-based client/server architecture makes it very convenient also for inter-organizational collaboration.

In pilot project Quickplace was used mainly for
- communication and collaboration inside consulting team as well as for collaboration "consulting team – small business"

- collaboration on preparation and approval of deliverables of project (reports, lists, Web presentations)
- project management – activity plans, tasks monitoring, calendars etc.
- knowledge repository (e.g. marketing database)
- (simple) workflow
- automatic notification of users by e-mail under defined conditions (e.g. weekly summary "what is new", new task)
- use of URL for access to other Internet resources (e.g. to project marketing database developed as Internet client/server system from this environment).
- marketing purposes - communication with potential customers
 - specifically for 1:1 marketing (personalized Web presentations for limited number of potential customers in step 5 of the method)

Quickplace server was used as service - in ASP

(Application Service Provider) mode. This has led to very short "lead time" of using it in project, as no installation and very limited administration activities were required. Quickplace's use in pilot project has led to considerable savings in project costs.

4. METHOD – DESCRIPTION

Method consists of the following 6 steps

- Step 1 - Introduction. Small business's analysis, Project – Start - up
- Step 2 - Business model documentation
- Step 3 - Selection of new (potential) markets
- Step 4 - Market research
- Step 5 - Internet marketing plan development
- Step 6 - Sale supported by ICT and I-commerce

Step 1 Introduction. Small business's analysis, Project – Start - up

Objectives

Main objective of this step is execution of small business analysis including firm description, organizational structure, financial situation, preparation of report on current firm situation, preparation of SWOT (Strengths Weaknesses Opportunities and Threads) analysis.

Second objective is mutual acquaintance of consulting team with people from small business. Consulting team introduces to small business concept of this method. Project manager (typically head of consulting team) presents to small business management the basic project plan (steps, dates, resources, deadlines) Project team members from both sides are appointed.

Key deliverables:

- document "description of the firm"
- document "SWOT analysis"
- document "project management and project team"

All deliverables (in all steps) must be approved by small business management. Decisions included in these deliverables are then interpreted as small business's management decisions taken on recommendation of consulting team.

Steps (Tasks):

- visit of consulting team to small business premises
- analysis of internal firm documents by consulting team
- interview of consulting team with management of small business (mainly in the areas of organization, production, sales, marketing)

Step 2 Business model documentation

Objectives

Main objective of this step is mapping and documentation of current "business model" mainly on the basis of internal company information, which can be – if the need be – extended from external sources. Consulting team acquires this information from small

business employees (mainly from sales and marketing) and eventually from company's ERP system, if available. Business model includes

- current products (properties, comparison with competition)
- production capacities (evaluation, possibility of expansion)
- current markets (description, CSF (Critical Success Factors))
- current distribution channels (partners, conditions, way of price specification, profit creation)
- current customers (products, volumes of sale, price specifications, customers requirements, customers satisfaction)
- specification of "core business" i.e. the main activity of the small business, which produces profit

Key deliverables:

- documents "products", "production capacities", "current markets", "currents distribution channels", "current customers". Some of these documents may already exist, some even in database form.

Steps (Tasks):

- analysis of firm's internal materials/documents by consulting team (see above)

Step 3 Selection of new (potential) markets

Objectives

Main objective of this step is preparation of list of several new potential markets (or their segments), which will be analyzed in greater detail in next step. List contains the basic attributes of markets (advantages/disadvantages for small business, entry barriers etc.) List should contain the markets, on which – according to current information from steps 1 and 2 – the small business could have a chance of success (i.e. realization of sale), i.e. such a markets for which there are no known barriers for small business's penetration of the market (product quality, prices, after sale service, market requirements etc.)

Key deliverables:

- list of new potential markets (including their basic characteristics)
- ev. basic vocabulary (terminology for foreign markets)

Steps (Tasks):

- searching in information sources (Internet, other digital sources – databases, traditional sources) in order to find new markets
- basic analysis of potential new markets (from the point of view of small business's penetration)
- small business management approves the markets recommended by consulting team

Step 4 Market research

Objectives

The main objective of this step is preparation of

cost-acceptable analysis of selected markets, taking into account that the result need not be as complex as results from much more expensive specialized marketing firms or from international marketing databases oriented towards area of specialization. Analysis contains information on markets and is used in the following steps of method.

Key deliverables:

- for each analyzed market
 - o market functionality (rules, entry requirements)
 - o geographic location (language, culture)
 - o volume (in prices)
 - o distribution channels (structure/model, most important players – leads)
 - o specification of prices
 - o market and channels participants
 - o potential partners, customers
 - o influencers
 - o competition
- marketing database including up-to-date and relevant marketing information on analyzed markets. It's content is updated continuously. It is used mainly in steps 5 and 6.

Steps (Tasks):

- analysis of individual markets (executed by consulting team with assistance of small business employees), using Internet resources and ICT with the goal of creation of marketing database

Step 5 Internet marketing plan development

Objectives

Planning of marketing activities, which will lead to interest in buying small business's products from new subjects in new markets. These subjects were identified in previous step. Although even the "traditional media" (phone, fax, print etc.) are used for these marketing activities, the emphasis of this step is on the use of ICT – namely Internet and Internet marketing.

"1:1" or "1:many" marketing as well as different ICT technologies can be proposed in the plan - E-mail, promotion on Internet sites, participation in vertical e-marketplaces, product presentation on Web-site etc.

Key deliverables:

- marketing plan

Steps (Tasks):

- preparation of marketing plan, which includes subjects of interest to small business as well as specification by which marketing approach they should be contacted

Step 6 Sale supported by ICT and I-commerce

1st sub-step

Objectives

Execution of marketing plan prepared in previous step (communication, providing potential customers the marketing/product information needed), which result in

expression of interest from these subjects.

Key deliverables

- marketing plan's execution

Final "deliverable" of this sub-step is expression of interest of potential buyers, providing them the required info on small business's products.

2nd sub-step

Main objective is to execute, especially by use of ICT and I-commerce, the remaining steps of commerce transaction – i.e. negotiation of contract (including price), delivery, payment, customer after-sales support. Small business uses also traditional means in these steps. But this sub-step puts emphasis on the use of ICT and I-commerce to supplement these traditional channels. However ICT and/or I-commerce can be used only if accepted by potential customer.

Key deliverables:

- execution of complete commerce transactions i.e. realization of sales (with potential customers), as it is also the main aim of all activities (both in methodology and in project)

Steps (Tasks):

- specification of means of ICT and I-commerce, which will be used for full execution of individual commerce transactions "partially" executed in 1st sub-step, with regard to specific conditions of each transaction.

As often negotiation of contract can take longer time (e.g. several months) the "final point" of this step in this project was set approximately to the start of two-side negotiation of contract on basis of information gathered during the project. Remaining activities were executed mainly by small business. If the need be, small business can use help of other subjects (lawyers etc) during the final stages of contract negotiations.

5. PROJECT

For project a production firm was selected, with about 400 employees, 3 product lines, limited marketing qualification, very basic brochureware Web-site.

Consulting team included 5 members. Project manager (located outside the country), 4 masters students (in different locations in the country) working on project part-time. They were oriented basically at - ICT (1), marketing (2), training (1). Project duration approximately 6 months.

Consulting team's activities - project management (including provision of professional Project Manager) and use of collaboration technology (Lotus Quickplace) were fully covered by IBM Software Group.

In Step 1 and 2 SWOT analysis was performed. Because of 3 different product lines, separate SWOT analysis

was executed for each product line. Project team was set-up for each product line, including 1 consultant and 2 small business employees. Quickplace was introduced and used heavily for communication and collaboration during the project, both inside consulting team as well as in "consultant – small business" area. Project organization, time-scale etc. was also set-up in Step 1.

In Step 3 - two or three new international markets were selected, based on consultant's recommendation, for each product line. Orientation towards international markets is important decision, as it influences many following activities, resulting in greater demands (on consultant, resources, costs, information etc.) than sole orientation to home markets.

In Step 4 several approaches have been used including Internet search engines, analysis of e-marketplaces. Between 20 and 70 potential new partners were located for each product line on markets selected in previous steps. Global marketing database Kompass was used on very limited scale, giving less than 10 contacts (for 2 Euro/profile).

Specific marketing database tool was developed in this step and used heavily from this step on. Database content was created in this step and used and made more accurate in next steps. Database contained about 270 entries.

In Step 5 marketing plan oriented mainly towards ICT and Internet's use was developed. Plan was oriented towards potential new partners located in step 4. "Value" of each partner for small business was evaluated (for each market and product line), which resulted also in order in which they will be contacted in next step.

As general approach 1:1 direct marketing was selected. For this purpose Web presentations in 3 languages were prepared in Quickplace. Collaboration technology is used in this case also for promotion. Individual presentations run in separate "rooms", which allow for great degree of personalization. Presentations are oriented towards small segment or in some cases even toward one (1:1 marketing) potential customer.

In step 6 the market plan was executed as planned, thus executing the "information step" of market transaction. Next transaction steps were executed mostly by tradition means, supported by e-mail.

Afterwards the results of the project were evaluated. The project achievements are shortly summarized below.

5.1 Project Results

Project's outcome can be – even taking into account some real-life limitations and shortcomings – evaluated

positively. Main project's goals have been attained

- method was developed
- small business has found new markets. At project conclusion there were 3 very promising negotiations under way regarding foreign markets which resulted in contracts later. One large contract has been signed in the home market. In this case the knowledge about the types of potential customers gained on international markets have led to contact in the home market with the type of partner which has not been contacted before.
- use of Internet –based ICT and Internet marketing for support of commercial activities has been demonstrated to small business and successfully used for gaining business benefits.

6. CONCLUSION

Based on results of practical verification the method can be evaluated as successful and as a useful tool for support of small business's commercial activities.

However implementation success is influenced by several success factors, some of which are basically critical (CSF – Critical Success Factors). As we have learnt, the basic CSF are the following:

CSF – global (in small business's country)

- the basic „support“ of small business's activities, i.e. state policies, legislation, taxes, governmental support
- this support strongly influences the way of project financing (which is critical). Financing can be provided by
 - o administration (state, region, Chamber of Commerce) – partly or entirely
 - o small business – partly or entirely
 - o sponsor i.e. ICT producer (as was the case of our project) or other – partly or entirely

CSF – small business

- fulfilment of basic requirements for expansion to new markets from production point of view, i.e. product which can compete with the products on international markets (in terms of quality, technology, price, after-sale service etc.), sufficient production capacity, ability to deliver product according to specific requirements of customers
- real interest in expansion to new markets
- strong support of small business management
- willingness of small business employees to take part in project activities and learn new approaches

- quick and qualified reactions to outside stimuli (e.g. reactions to questions and requirements of potential customers/partners)
- project effort and it's length of time depends on level of knowledge of small business employees especially in areas concerned (marketing, ICT, foreign languages for new markets)

CSF – consulting team

- good project management
- qualification of team's members and their „entrepreneurial“ approach
- language qualification (for new markets) or possibility of external provision (for not so common languages)
- team spirit

Based also on results of this case our University has launched an open „Program of cooperation of Universities and Small Businesses“ called eMSP. Program is opened to any University in the Czech Republic, small businesses, administration, industry – producers, suppliers and offers services oriented towards support of various small business's activities.

7. REFERENCES

- [1] Chaffey, D., Mayer, R., Johnston, K., Chadwick, I., *Internet Marketing*, Essex: Pearson Education Ltd., 2000
- [2] Hanson W.: *Principles of Internet Marketing*, Cincinnati: South-Western College Publishing, 2000
- [3] Hejtmánek, J., „Use of ICT at SME for penetration of new markets“ Diploma theses, University of Economics Prague, Faculty of Informatics and Statistics, Prague, Czech republic, 2003, (in Czech)
- [4] Hoffman D.L., Novak T.P., Chatterjee P., „Commercial Scenarios for the Web: Opportunities and Challenges“ *Journal of Computer-mediated Communication*, Special Issue on Electronic Commerce, Vol. 1, No. 3, available at <http://elab.vanderbilt.edu/about/faculty/tnvita.html>, (visited 16.6.2004)
- [5] Jandos, J., „IT Market in Czech Republic and It's Support by Internet Commerce“ *Proceedings of BITWorld 2001 Conference*, pp. 186 - 196, Cairo, May 2001
- [6] Kalakota R., Robinson M.: *E-business, Roadmap for Success*, Reading: Addison Wesley Longman Inc., 1999
- [7] Munkvold B.E., *Implementing Collaboration Technologies in Industry*, London: Springer Verlag, 2003
- [8] O'Connor J., Galvin E., Evans M.: *Electronic Marketing, Theory and Practice for the Twenty-first Century*, Essex: Pearson Education Ltd., 2004
- [9] Rayport J.F., Jaworski B.J., *Introduction to E-commerce*, New York: McGraw-Hill Higher Education, 2001

The Design of a Lean CRM Software Process Model

Chen Bocheng¹ Ip W. H.² Zhou Yuebo³, Liang Bing², Yu H.L.²

¹ School of Economics & Management, Tsinghua University, Beijing100084, China

² Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, China

³ SAP company, Beijing 100020, China

{chenbch, liangb}@em.tsinghua.edu.cn

ABSTRACT

In this paper, the concept and development of CRM (customer relationship management) are discussed, we defined a customer "lead" concept which emphasized the inter-connection of downstream and upstream processes of our customers. It is pointed out that the key point in CRM is the customer lead management in the software design and implementation. Based on this perspective, the processes of lead processing, lead management, and inside lead management are further elaborated, analyzed and their models are proposed. The problem of analyzing customers' behavior is also studied. Hence a simplified and lean integrated CRM software process is finally given, and this integrated model can be used to build CRM software for small and medium-sized manufacturing enterprises.

Keywords: CRM, Customer Relationship, Lean Model, Business Process Management

1. INTRODUCTION

Discussions on customer relationship management (CRM) have been going on for several years. International studies have tended to focus on the management function of CRM and its management characteristics^{[1][2]}, and less on analyzing the structure and processes of such a system. We rarely see more in-depth research or thoughts on system design in the area, although CRM solutions have been developed by many software companies.

Referring to SAP company's CRM materials^[3], online help document^[4], and the Siebel company's Demo software^[5], we study and analyze the structure of CRM software. This paper emphasizes the re-construction of the CRM concept and provides a basic process framework of CRM model suitable for analyzing small and medium-sized manufacturing enterprises (SME). A typical manufacturing enterprise means a company has post-sales service and permits customers to configure the product they need. The basic process means a general framework, not taking industry characters and more complex processes into account. Generally, CRM software consists of three levels: the operating level, the level of analysis, and the level of cooperation. The model discussed here contains only an operating level, with little analytical function.

The object of building a lean, simplified model is to integrate the functions of marketing, sales, service, and configuration in an enterprise. Enterprise marketing and sales management will then be promoted from the departmental to the enterprise level. Another object is to make CRM software easier to understand, because managers may find it difficult to understand how to operate an enterprise supported by modern IT technology, which may lead to great changes in the relationship an enterprise has with its customers.

It should be pointed out that CRM software is just a software tool. What this paper discusses is the software design, development and implementation processes and how to improve them to help manage business flows in an enterprise.

2. MODEL STRUCTURE ANALYSIS

The first difficulty in modeling is how to describe the "customer" and "customer relationship." The words "focused with customer" can be seen in every paper on CRM, but left un-resolved are the questions of what a customer is, what a customer relationship is, how such a relationship is to be made the focus in the software.

Contact Center	Phone	Internet	Direct Contact
Marketing activity / Marketing Department	Tele-Marketing	Internet marketing	
Sales activity /Sales Department	Tele-Sales	Internet Sales	Mobile Sales
Service activity / Service department	Service Interaction Center	Internet customer self service	Mobile service

Fig1. Contact Center Structure

Also less discussed is how to build the process structure in CRM software to realize the concept of "centering." We see many variations software, but there are fewer papers describing why they are designed to look they way they do. In the following, we attempt to conduct such analysis of the CRM software. "Centered" should mean the monitoring, management and response to the feedback from the customer, i.e., the market. First, enterprises should be able to monitor the feedback from each "connection" between departments and the customer. The feedback is obtain by the connection, and the relationship is also from the connection, so the model must be related to the connect center. The model in the paper is a 3×3 connect center (Fig. 1) that covers almost all functions and various corresponding methods of connecting. Though there

may be only part of them in a real enterprise, it will not affect the discussion behind. The system should be able to manage the incoming customer information and send

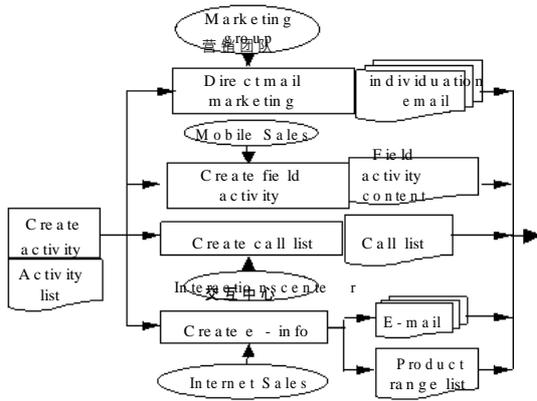


Fig 2. Sending clue process in the Marketing level

it quickly to the corresponding stages, as well as give a response.

2.1 The Clue and Character

A careful analysis of the function of the connection center, which has a 3x3 connecting function (in all orientations) with customers, reveals that all of the results of its operations (on information flow) can be abstracted in one word – “clue.” Each piece of incoming information has its own source, and each piece of information that is sent has its own destination; they connect customers just like a invisible cobweb, regardless of whether or not a sales order is made. The process of forming a relationship between the customer and the enterprise is: connecting - understanding - making a relationship - exchanging action, so the clue is what we want to find – the basis of the customer “relationship.”

The reason why we emphasis the concept of the “clue” is also because it supplies us with a new perspective from which to rethink the design of CRM software.

Let us re-examine the concept of customer relationship management. Although not defined exactly, the “relationship” is, in fact, related to information on the customer and to the current and past connections between enterprise and its customers. Therefore, to some extent, we can say: to CRM software, customer relationship management is the processing and management to customer “clue”. Centered with customer, to the model, means centered with connecting, with clues, to complete monitoring, and replying the customer information in real time. This point should be the center of the design of the CRM process model; its software management process design should address this point. The accumulation of clues is also one resource of an enterprise, and the management of such clues promotes the management of the relationship between the enterprise and customer from the level of the department to the level of the enterprise. For the

same reason, an analysis of the enterprise’s relationship with its customers should not only focus on those with an order history, but on all customers who used to relate to the enterprise. This is because the aim of marketing is to enlarge the customer group.

Clues can be divided into three parts by the time of connection: connecting now – current clues; connecting in history – database clues; marketing connecting, aimed at enlarging the number of sales orders, and customer group based on customer classification and taste (target group), which are obtained by computations based on current and historical data to yield future clues.

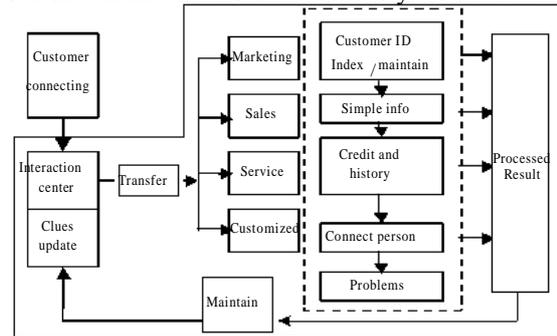


Fig3. Clue Receiving Process

The clues can also be divided into two parts based on inputs and outputs to the software system:

The first is sending clues. This part also covers two parts: answers to daily business processing; and marketing action information to obtain more customers. The process, which an enterprise needs to maintain and enlarge its market share, is illustrated in Fig 2.

The second part consists of coming clues. It also includes two parts: inquiries about daily business products or services (information less related to sales orders); and receiving clues, which are information about sales orders or plans for sales orders. The process is illustrated in Fig. 3, where the process in the dashed line block indicates the parts shared among several departments.

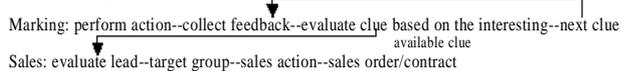
2.2 The Process to Manage Clues

The clues can be divided into available clues and general clues (not taking garbage clues into account). The available clues refer to the clues thought to have been changed to sales orders with a larger possibility, and the rest are general clues. But general clues yesterday may be available clues today.

2.2.1 The Process of Dealing with Clues

The processing of clues by CRM software may be divided into two levels: marketing level processing and sales level processing. We can then obtain two methods: 1) Medium-sized enterprises can first deal with the clues at the marketing level, and determine the clues that are available for transferring to the sales level. The sales

level then determines the clues that are most likely to lead to sales orders and to one-to-one services (two-level processing). 2) Small enterprise always need one level of processing, i.e., sales-level processing. As to the word “clues,” Siebel’s Demo refers to it as opportunity, SAP refers to the clues available in market level as lead, and to clues available at the sales level as opportunities. We define clue processing process based on the two-level process of clues. The evaluation process in such a system (divided market and sales) can be shown as follows: (the perform action in the marketing level can be seen in fig 2)



Therefore, in the process from marketing to feed-back, and from feedback to sales order, all exchanges are arranged to be completed from the connect center. If managed better, the clues in the process can all be recorded.

2.2.2 The Customer Characters in Clues

The customers of an enterprise always mean those with a history of commercial exchanges with that enterprise, but the concept of the customer in CRM is rather different. Fig. 4 shows the different concepts of the

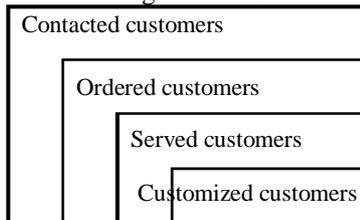


Fig4. Inner-relation among different customers

customer and the inner relationship among them: the connecting customer is the total set of clue: the order customer means those with a sales relationship; service

customer and customizing customer are obviously sub-sets of order customer. Enterprises can select interesting groups of customers from different levels in Fig. 4 when they want to classify customers, but on the other hand, this will also show at least some of their attitude towards operations and the marketing environment, allowing enterprises to select subject products and target groups from various customer levels.

- z Pessimistic attitude: selecting a target group from the order customers, with the process of sending clues aimed at those customers. The object is to first fix upon the customers, then enlarge the market share. At the basis of the theory is the belief that the cost of developing a new customer is five times the cost of maintaining an old customer, and that 20% of the customers produce 80% of the profits of an enterprise. Such method is always adopted in a bad market environment. maintaining a old customer, 20% customers produce 80% profit of enterprise. Such method is always adopted on the bad market environment.
- z General attitude: selecting a target group from the order customers, with the group of clue customers also the same as that of the connecting customers.

The aim is to enlarge customer term furthermore at the time of fixing its own customer term. The advantage is that an enterprise can get extra customers, besides their original customers, and need not worry about omitting the requirements of the original customers.

- z The optimistic attitude: selecting a target group from the connecting customers, where in the sending of clues, the customer group is also the connecting group. Its object is to enlarge market share and the definition of the customer. In better market conditions, such a method will generally be adopted. The defect is that the target group obtained from clue customer may be some different from that obtained from order customers, or may even cover the requirements of the original customers.

2.2.3 The Managing Process To Clues

The model’s Object aims to integrate enterprise’s marketing, sales, service, configuration and analysis method into one frame. From SAP’s training material^[3], the clue can be managed by following process. First following concept should be given at least in the model.

- z Marketing- to enlarge market, it is decided by enterprise stratagem, operating object and it own resource and capability, and its document is market planning.
- z Campaign- to develop a market needs several campaigns, each organized campaign aims to a special subject product and target group, those customer group is got by computing in special arithmetic, and refer to the outside analysis information and the third analysis result, its document is campaign plan;
- z Action- a campaign is composed of a series of actions, and its document is action content.
- z Subject product- one kind or a series of product planned to push up (usually are higher ROI product);

Target group- a kind of classified customer got by computing the interesting of connecting clue customer and history customer data, its document is customer list;

Their relating is shown in Fig 5. The planned action is performed in sending clue process of marketing part(fig 2), when the campaign performed. And each action should be affirmed in all sending process to make sure it is performed. The market and campaign can be combined into one level clue management in some small system case. The work that needs to be performed after sending is the collection and analysis of clues on feedback.

2.2.4 The Inside Process of Managing Clues

From the literature review of SAP material^[3] and Siebel’s Demo^[5], the inside management of clues can be adopted following a proper structure. The following

information are required to be in the CRM database:

- z Clue—a clue code is a directly connecting code for customer by various of channel and connecting media such as call numbers, mobile numbers, email addresses, and so on, so that the system can find the customer code using each clue;
- z Customer code — may directly index the customer’s enterprise information, enterprise account, and connecting history;
- z Connecting person — there may be several connecting persons in an enterprise, who may be responsible for different businesses or for the same business; the system can find a list of connecting persons by customer code, in case this is needed.

When a connection occurs (taking call numbers as an example), the system should list those customer enterprises that have a similar call number as the coming number, and allow the receiver of the call to read messages related to the incoming call, such as the company code, simple information about the company, the company’s credit status, and come of its commercial or connecting history. Furthermore, should allow him or her to read those records about the connecting people before and as part of the process of exchanging information (the clues involved are in the clue database illustrated in Fig. 5; its process is illustrated in Fig. 3). The advantage of this is that a business person (the receiver of the call) in same role can quickly know with whom and with which company he is communicating, and know its connecting history and the level of its relationship with his own company. From the simple record, he can anticipate what will be discussed, what questions will be asked, where the difficulties may lie, and the solutions previously given, and so on. Thus, the problem can be quickly affirmed and a new solution will be given.

2.3 A Comparison with ERP

Based on the above discussion, we can make simple comparisons between the ERP and CRM software.

- 1) The difference in aim: ERP is software that is based on process and centered on product management; CRM is based on process, and centered on clue management.
- 2) The difference in customer concept: customers in ERP always mean those who have an order history-certain customers; customers in CRM mean clue customers who used to connect to the enterprise-uncertain customers. There is more information on future customers in CRM.
- 3) The difference in information processed: the information processed in an ERP system is information that is currently useful (a series process after the making of an order), while the information in a CRM system is mostly currently useless, or useless forever, but no one knows which part may be useful and which part may be useless.
- 4) The difference in input/output: in an ERP system, orders are the input and products are the output; in

a CRM system, clues are the input and sales orders are the output. So the interface between ERP software and CRM software should be the sales order (as defined earlier).

- 5) The difference in model: the ERP model shows the integration of the whole enterprise process; and the CRM model shows the integration of the whole market and the sales process.
- 6) The difference in reference. The ERP model addresses enterprise business management framework supported by IT; while the CRM model addresses to the market and sales business management framework supported by IT.

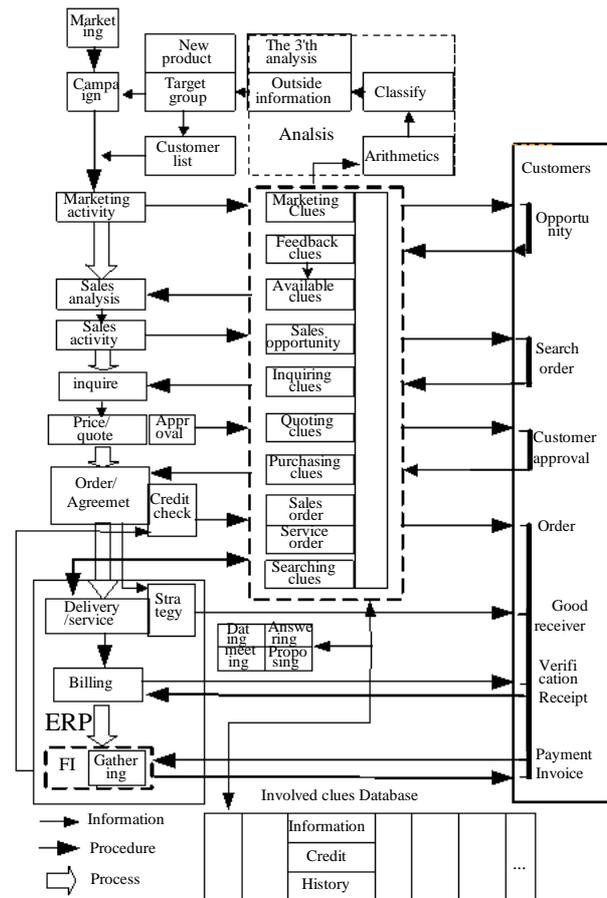


Fig5 CRM software simplified process model

3. THE INTEGRATED PROCESS MODEL AND ITS APPLICATIONS

A lean simplified, integrated process model that summarizes the above analysis is shown in Fig. 5. The model takes into account the enterprise functions of marketing, sales, services, customer configuration, and analysis, in order to make a CRM integrated process framework designed from the market orientation of the enterprises and customer orders. In Fig. 5, the thick line represents the process, the thin line represents information, and the dashed line box represents connecting center and its corresponding receiving and sending processes. Besides the marketing sending process, the marketing part also has the function of dealing with

general clues. Its result may be a simple answer: the giving of a suggestion, fixing a time for a face-to-face talk, or arranging a small meeting. The total process in Fig. 5 is as follows.

When planning a market activity, the people making the analysis begin from the clue database, refer to outside information and obtain the subject product and classified target group (customer group). They then come up with the object of the campaign, and organize promotions to reach the designated group of customers. In order to attain the objectives of the campaign, the market department should organize a series of corresponding actions, design corresponding action content, and perform them through the connecting center. As the campaign proceeds in all directions, in this target group old customers gain a new understanding and new customers who become interested in the subject product may connect to the enterprise's marketing department. The marketing department analyzes those clues on feedback, selects available ones, and transfers them to sales department. Those making inquiries are transferred directly to the sales department. The sales department studies this information in order to determine its own target group, then organizes corresponding actions in order to turn opportunities into sales orders.

The sales department also has the problem of receiving feedback. After orders arrive, CRM transfers the information to ERP. The ERP system completes the translation from orders to products and then to corresponding function. CRM takes the corresponding records, in case the customer inquires about it. Of course, the system should have the ability to compute costs, to evaluate the market and the effect of the campaigns. Obviously, in this the model, CRM cannot complete the order independently, so this part should have an interface to the ERP. The interface is not just a data interface, but needs to ensure process and organization, otherwise the system can only transfer data, but not complete the business processes. In fact, we can refer to the SAP's mapping method between SD module organization structure and that of the CRM^[3,4].

The sales part in CRM should have some similarity with that in the ERP. For example, the processing to inquiring/quotation, processing to order and so on: customer inquiring-enterprise quotation refers to the commercial history- customer making PO (purchasing order)- sales people searching inventory- enterprises sending SOs (sales orders) - the sales order to ERP system. The order is the output of the CRM software, and the later one is in the ERP manufacturing process. The service part has its own process, and is also processed in the flow of information, until the SO's are sent out. Service orders in a manufacture industry are

always related to manufacture orders: service order processing-resource planning-in site service completion and affirmed -- billing -- account receivable.

4. CONCLUSION

A lean, simplified process model of CRM software has been proposed, and an analysis of CRM using a "clue" concept has been given in this paper. Our greatest contribution would be the concept of the "clue": for the database built based on connections is the basis upon which an enterprises analyzes customer behavior and characteristics, and is also the basis used to integrate and predict, and establish customer "relationship management." In today's ERP/MRP/II software which is based on process and centered on products., it is essential for monitoring performance in real time, reacting in time to the status of the market, customer status, and customer feedback to form a CRM core management framework using the process and centered on customers (clues). We can see that CRM concepts, methods, and meanings may change with time, technology, and customer requirements, or the changes may even introduce a change in the structure of the CRM model, but these changes can be accommodated with a lean and simplified structure described in this research.

ACKNOWLEDGEMENT

This study was partly supported by the National Natural Science Foundation of China (70231010).

REFERENCES

- [1] Winer, Russell S, A Framework for Customer Relationship Management[J], California Management Review, 43(4), 2001: p89-106
- [2] Ling, Raymond; Yen, David C, Customer Relationship Management: An Analysis Framework And Implementation Strategies[J], Journal of Computer Information Systems, 41(3), 2001: p82-98
- [3] SAP company,SAP CRM training material[R],SAP company, 2001
- [4] SAP company,SAP online help[OL],http://help.sap.com/saphelp_crm30/helpdata,2002.9
- [5] AMT site,Siebel CRM software demo 6.0[OL], <http://www.amteam.org,2001.12>
- [6] Scott Nelson, Tom Berg, Customer Relationship Management: An Overview[R], Gartner Group Strategic Analysis Report, 2000.8

The Development of RFID in Healthcare in Taiwan

Frank Kuo¹, Yuehching Lee¹, Chuanyi Tang²

¹Institute of Healthcare Information Management, National Chung Cheng University, Taiwan, China

²Institute of Information Systems and Applications, National Tsing Hua University, Taiwan, China
frank.cv85@nctu.edu.tw

ABSTRACT

Healthcare industry has unique characteristic about human's life. Any mistake or medical error will cause the irretrievable regret. Patient safety is the most important issue in recent years. Because the uncertainty of disease occur, the demand of healthcare need, location and types are changed frequently. By the pressure of "Self Management Plan", how to make limited healthcare more effectiveness and efficiency is critical for survival. Radio Frequency Identification (RFID) can be able to provide to identify, tracking and tracing for patient or medical objects directly and continuously. RFID is made object wisdom and process automation. It debase the cost of collect data from first-line in tradition MIS.

RFID in Taiwan is more famous. There are four kinds of applications in healthcare industry. Some of them will be the killer application when government's policy trigger and industry full-scale implement in the future. There are six hospitals deploy RFID in pilot plans by government support. And there are still some necks of bottle in deployment. The research assisted by foreign research and extracts by the experience in expert interview. We point out the critical deployment factors by five dimensions, including: system technology (system readable rate, frequency limitation, instrument interference & Human healthy, and transmission distance), system cost, common standard, health industry infrastructure and system integrate, and the issues in privacy and security by patient. The future research tries to find out the total solution to deploy RFID in healthcare industry to build up the "patient safety healthcare environment".

Keywords: Radio Frequency Identification (RFID), patient safety, healthcare industry, medical error, bar code

1 INTRODUCTION

Healthcare industry is related to the person's life closely. Any mistake or medical error will cause the irretrievable regret. Longer work time and work pressure made medical errors and negligence often happened in Hospitals. Healthcare is a high human concentrated industry. There are a lot of team works and coordination together from emergency medicine, outpatient care and inpatient care in hospitals. The report of Institute of Medicine (IOM) in "To Err is Human-building a safer health system" indicates that medical error often happens on bad communication and artificial negligence [6].

Patient safety is the most important issue from World Health Organization (WHO), European, America and other nations in recent years. After experiencing the storm of SARS, beating the wrong needle and giving the wrong drug...etc. The healthy quality and patient safety subjects are more valued by the public gradually. "Department of Health" considerably pays more attention to patient safety, and announces five execute goals in annual plans for government policy in 2004 [2].

Healthcare industry has his unique characteristic. Hospital is a place that people care of people. The healthcare service can't save and shift. Because the uncertainty of disease occur, the demand of healthcare need, location and types are change frequently, so the maximum difference between healthcare and others

industry is the uncertainty of medical demands [5]. There is high inside knowledge and entry barrier in healthcare industry. Computerized systems have capacity to augment the physician's capabilities to reduce the error rate for the advantage over humans, include: never forget, never get tired, and make no mistakes [8].

Hospitals in Taiwan had long history and abundant experience to import Hospital Information System (HIS) and Electronic Healthcare Record (EHR) with IT. But, there are more competitive and high pressures in hospitals in Taiwan by the limited income for healthcare insurance in "Total amount payment method", "Self Management Plan". How to deploy information technology IT to assist specialists in hospital working more effectiveness and efficiency to increase capability in limited medical resource and control operation cost cautious are the most famous research issues to the scholars in healthcare information management and healthcare service management [1].

Now, HIS can purely support work for operation level or management level. But, the most bottlenecks to implement information system in hospitals is that HIS can't handle the information about object name, time and location correctly including medical workers, patients and pharmaceuticals. Then it's hard to advance the capability for HIS to statistics, analysis and decision support, even more than strategy level.

Radio Frequency Identification (RFID) is one of the top

ten technologies in this century [14]. RFID is a kind of wireless technology by using radio frequency (RF) to identify objects. RFID can be able to provide to identify, tracking and tracing for patient or medical objects directly and continuously [1]. The characteristic of RFID is made object wisdom and process automation. RFID can reduce the cost of collect data from first-line in tradition MIS and have dramatic potential to great efficiency and profit in supply chain management (SCM) from manufacturer, logistics to retailer [4].

RFID can decrease medical errors, promote healthcare quality, and make the healthcare process automatic and optimize to save amounts of work in data collection, entry, integrity and confirmation such as patient identification, the three reads and five checks when giving pharmaceuticals, the data exchanged by patient referral...etc. By the characteristic of PPI (Positive Patient Identification) and POC (Point of Care)[16], RFID technology makes the maximum effect on HIS to deliver information at right time, real time and any time by reduce the time in data entry, object identify and make healthcare process visibility.

RFID is not a new technology, but it's a new revolution for application. By the same standard, protocol and communication language, there will be a strong impact for our life. Like the evolution for geographic commerce, electronic commerce (EC), mobile commerce to ultimate commerce, there's also extend to "U-healthcare"[21]. There are 4U in healthcare which includes: "Ubiquity" healthcare and service, "Uniqueness"EHR, "Unison"information and patient integration, and "Universality"healthcare environment. RFID provides "Competitive Intelligence"to decision maker and makes "Quick Response (QR)" for healthcare market demand.

2 RFID IN TAIWAN

RFID in Taiwan is more and more famous. Taiwan government provides both reward and investment by "Ministry of Economic Affairs" (MOEA). The substratum of administrations in MOEA are "Department of Commerce", "Department of Industry Technology", and "Industrial Development Bureau"[30]. The direction of policy in "Council for Economic Planning and Development to president" and the conclusion in "Science and technology Advisory Group of Executive Yuan" in "2003 SRB (Strategic Review Board) meetings" also promote RFID to an advance and significant development in Taiwan[23].

The "Executive Yuan" (the Cabinet) in Taiwan promulgated "The vision in Taiwan of strategy service industry in 2010". MOEA established a cross-division RFID promotion team. The three important guide lines are "The Alliance of Supply Chain, (S plan) ", "The research and application in RFID platform, (R plan)", and "Telemetry, Health tourism, Integrated

medical system in Healthcare , (THIS plan)"[23].

There are different RFID industry orientation in different core development fields. RFID industry can discriminate in five divisions: RFID Tag (chip/die, antenna, attach, testing, convert), Reader (RF circuit design, digital circuit design, API development, and antenna module), Middleware (like SUN), System Integration (like IBM), Application Software (like Oracle), and Consultant (like Auto-ID lab).

Taiwan has maturely knowledge for RFID in semiconductor design, memory design, and specialized wafer production, electronic fabrication, software development, and system integration. Wireless and communication industry provide the capability of RF design, antenna design and field testing. Automatic equipment industry also possesses packaging and testing ability to cost down and increase production [19].

"Executive Yuan" formed RFID technology research and industry application alliance include six special interest groups (SIGs) in Taiwan, including with: Pcess equipment and Mterial SIG, Design and Mnuufacturing SIG, System Integration SIG, Supply Chain SIG, Testing and Certification SIG, and Industry Information SIG [32].

Many foreign information corporations also build up RFID research centers in Taiwan by the "International Innovation and R&D Base Plan" in "Challenge 2008-National Development plan"[22]. The enterprises include Oracle, SUN, HP, and Microsoft... etc. One of Auto-ID labs is in Fudan University, Shanghai. Two sites cross-strait of China are the main base of manufacturing center for foreign corporations. There will be many RFID technology developments and success application to paradigm shifts both in china cross-strait in nearly future.

3 RFID IN HEALTHCARE

3.1 The trigger of RFID in healthcare

There are no issues more important and popular than patient safety in Taiwan or foreign abroad. HIMSS patient safety survey indicates bar codes and CPOE (Computerized Provider Order Entry) have potential to improve patient safety [15]. Frank Kuo indicate five applications of RFID to increase patient safety by the famous seven patient safety goals in JACHO 2004, which including blood management, drugs management, surgical operation management, infection management, and medical equipment management [2]. RFID can also provide functions to satisfy three main demand of FDA's 21st century approach to patient safety like automatic data collection; partnerships with other entities (government agencies, healthcare providers, and payer organizations); and enhanced communication[26]. Wal-Mart requests the providers of front hundred

pharmaceuticals must deploy RFID in package with EPC standard, Read-Write passive tag (class2) before January 2005[27]. Accenture Technology Lab makes up the alliance of the biggest of eleven pharmaceutical manufacturing, distributions and hospitals [9]. Healthcare Distribution Management Association (HDMA) who constructs the health products and distribution standard to order that every pharmaceutical must paste RFID tag by case level of pharmaceutical in 2007[33]. The regulations of the "Food and Drug Administration (FDA)" in USA are even stricter to item level [27].

The specially report of "ID TechEx" estimate the application market of healthcare in America medical industry in 2010 will be arrive at \$ 86.3 billion[29]. By the above of the pharmaceutical supply chain with RFID, it's a trigger to detonate a great quantity of the downstream hospital implement RFID. So the questions of hospital aren't either or not to deploy with RFID, that is when to do it? [20]

3.2 The development in healthcare in Taiwan

The development of RFID in Healthcare in Taiwan is just at the initiation stage. There are six hospitals proceeding to RFID pilot plans earlier by government support. The pilot development goals of them are patient identification, drug identification, contact history tracking, infection announce, and medical warehouse management [1].

During SARS period, "Ton Yen General Hospital" deployed RFID pilot plan which technique support by "Industrial Technology Research Institute (ITRI)" at the first. And then, "Taipei Veteran's General Hospital", "Taipei Medical University Hospital", "KFAYA cancer center", "Show-Chwan Healthcare System", and "Tri-Service General Hospital" proceed RFID plans by government support[34,35,36,37,38,39].

"The Promotional Committee of e-Business Standard in the Medical Industry" builds "the Common Trade Platform in the Medical Industry" by "E-Commerce counsel plan" in MOEA. Taiwan wants building pharmaceutical data exchange standard and environment by XML and web service to promote RSS (Reduced Space Symbology) and CS (Composite Symbology). The committee will try to integrate EAN.UCC standard to healthcare industry [25].

The development directions and application fields with RFID in healthcare can division into four parts:

(1) Biotechnology and Pharmaceutical Fields:

Chemical reagent tracking, Animal, cell, and patient tracking and compliance in clinical trails.

(2) Medical Supply Chain:

Pharmaceutical supply chain, Warehouse management,

Anti-counterfeiting, Blood bank management, and Monitoring radioactive isotopes. (Including drug and surgical dressings)

(3) Hospital Internal Management:

Patient identification and position, Drug compliance, Disease control, Coordination the medical device, and Clinical path optimization.

(4) Healthcare Service:

Disease control, Emergency medicine, Sensors and Telemedicine, Long-term care, and Community continuity care.

4 THE EXPERT INTERVIEW

RFID implement in healthcare still in initial stage. There s no ready case or completely success experience in healthcare. We proceeded depth expert interview by quality research method with 10 experts, which including chief of the hospital information department, and scholars with RFID research in healthcare information, and health service management in Taiwan. We find out the advantage and inferiority in RFID in healthcare.

After expert interview, we integrate, sort and analysis the view and opinions in expert. The mainly critical key point to deploy RFID in healthcare industry can be classified into five followings: System technology, implement cost, standard protocol, information infrastructure, and privacy protection. We discuss them as follows:

4.1 System technology

(1) System readable rate

The most special features of RFID system lies in non-line of sight. Radio frequency delivering in air-interface needs better channel coding and modulation to reduce the signal broken and noise. Currently the readable rate of RFID system can't reach 100%, and this also becomes a biggest obstacle to implement. There are a lot of primary factors which influence the system readable rate, including: the direction of transmission, air-interface, the characteristics of product, and the characteristics of operation (environment).

The electrical label and reader have the directivity so there is extreme difference on the reading angle (horizon, vertical, oblique). If the product was horizontal between each other and outward to the product, it could be improving the readable rate. In other words, these critical factors such as temperature, humidity, and electromagnetic interference (EMI) of external environment should be paid a lot of attention. On the other hand, if the electrical label is put on the counter of absorbent or metal that would be decreased the readable rate substantially. The other reasons that

make the low rate of reading such as the distance of transmission and the moving speed of tag.

Besides, distance, indoor environment, the design of antenna, and the direction of installing are also very important influence factors. We can't use the high power reader in hospitals. Therefore, while implementing system, they should depend on the electric wave strength, clinical path, the need of moving line, and the result on-the-spot test which make they decide the demand and the position of readers. So, to deploy RFID isn't only a technology, it's an art and experience.

(2) Frequency limitation

The wireless channel belongs to the public wealth. The International Telecommunication Union (ITU) is all carried on the programming and allotment of the frequency application by the government supervisor organization. Using RFID hardware equipment needs passing license by the "Ministry of Transportation & Communications" in "The Directorate General of Telecommunications (DGT)".

There are open frequency in the industry, about scientific and medical (ISM) in 13.56MHz, 2.45GHz. Currently, 915MHz is a controlled radio frequency, but the Directorate General of Telecommunications (DGT) has already declared the end of the year will open the 922-928 MHz channel to provide the RFID usage in the end of the year [16].

(3) Instrument interference & Human healthy

It is the most anxious reason that radio wave interferes of medical instruments interactive, and it also influences weak patient in the medical environment, too. The conclusion of FCC and IEEE didn't find out any evidence that "Non-Ionization Radiation" of the normal regulations usage will endanger the human body. But nobody and no research guarantee about it.

On the interference of the medical instruments, we need to advance the demonstration first before implementing RFID in hospitals. Then, we carry on ducting into the work and field-test. Hospitals can aim at first: Electric shock machines, electrocardiogram (EKG), breath machines, and IV Pumps etc equipment carries on the test, and measures its wireless delivers the data mistake rate (BER) and electromagnetism mutual interference (EMI) condition. Currently hospitals which have implementing RFID all did not find obvious influence. However, the telecommuni- cation bureau and department of Health supervisor organization should ensure safety to relieve the misgiving by more professional research [12].

(4) Transfer distance

The factors which influence transfer distance include: reader power, sensitive of RFID tag, and four factors above-mentioned that influence about system readable rate. There are two kinds of RFID tag type is passive

and active. Active has battery by itself, but passive hasn't. Under the common testing environment, the distance of passive tag of 13.56MHz is approximately 1m, and 915MHz (UHF) is about 5m; the distance of active tag that the 2.45GHz (Microwave) is about 3m. The active RFID tag with 915MHz can transfer close to 100m.

4.2 Implement cost

Recently, there are some developing hamper in the other industry, main reason is higher cost. Healthcare has strong relation with persons' life and any mistake or medical errors may cause the irretrievable regret to patients. Medical errors will be impact hospital's brand and decrease customer's confidence violently. Furthermore, Healthcare should belong to public-spirited wreath in nations, it should be considering how to value added in society, not just to earn money. Medical institutions need to promote medical quality, decrease medical errors, and raise working efficiency with sacred spirit. The potential benefit and reality must evaluate altogether.

But RFID tag cost still influence spread tempo. The method of production and the market scale of RFID are the key factors to cost down. Just like the noted question about "Chicken birth eggs first, or contrariwise egg result chicken." Auto-ID estimate at the threshold limit price is \$5 cents. But, the price of purchasing in market is still around 40 cents (or above). Taiwan government announces that "Quanta Computer" can already production RFID tag around \$5 cents [11].

Like the statement to implement EMR, and HIS before, the cost of maintain always be extreme high than initial deployment expense. When hospitals evaluate to deployment RFID, they should be realized the capability, function, budget, and characteristics of their demand. To ask for advice with consultation is greatly important.

4.3 Standard protocol

There is long history in RFID, and many nations had already established their standard specifications for RFID, such as ITS America, U.S. Department of Defense, AIDC, ISO, ANSI, and CEN...etc. Because EPC network has the special feature such as automatic data exchanging, it extensively applied in general so far. The standards such as the computer network infrastructures of EAN, UCC, and MIT Auto ID Center propose EPC Network. The standard of EPCglobal includes the coding structure on tag, air-interface, local interface, Savant, Physical Markup Language and the Object Naming Service (ONS) together with the Savant support network [24].

In the aspects of foreign medical industry, Health Industry Business Communications Council (HIBCC) establish common standard protocol of the medical care

environment and the protocol can communicate with ANSI, CEN, and HIBC[28]. Currently the domestic medical hospital has no common standard. However common coding format, it appears that data interchange among different HIS is more important at medical institutions. This needs the Department of Health and related supervisor organization to establish it as soon as possible.

4.4 Information infrastructure

The complete infrastructures is a cornerston to deploy RFID well, including : the network communication with common protocol, Health Information Network (HIN), PML server, ONS server, and the mechanism about audit and authentication, and un-interpolation. On the other hand, the backup mechanism should be established beforehand. Through the high utilization, stability, scalability of server system that can maintain a lot of data file, analysis and process.

By system integration view, the simple close loop system (single hospital inside) can be extended the open loop system (inter-hospitals by alliance). But it will substantially increase system complexity, expenses and timing significantly. However, it's only amplified on more effectively and reach the maximum revenue of investment (ROI) in implementing RFID, when the information infrastructure has been build up completely.

4.5 Privacy protection

The right of privacy here can be divided into medical workers and patients. Medical institutions can control more complete and durable information of members in hospital by RFID, but also cause larger personal controversy of privacy. Medical industry has its particularity. Under the need of safeguarding the whole benefit of social, the public hygiene environment and medical science research etc, personal privacy is not placed into the initiative of position definitely while medical information privacy and personal right conflict mutually.

Besides, the questions in RFID about EMR data ownership, data security responsibility, data expiry date... etc, they also cause many controversies between medical care quality and the power of professional medical treatment from academy such as medical science, law, politics, science, technology and humanities etc.[3] The supervisor authorities should carry on the more thorough study about this, and have definite standard to let medical institutions to follow.

5 CONCLUSION

Like the chairman Mao Tse-tung in China who said: "The future is glory, but the road is complicated". And he also said: "Doing right direction, everything will be alright"[40]. To understand the development process

and constrains of RFID technology in healthcare, can let us deploy RFID correctly at right time, right place and right strategy model. Then, hospitals can possess best ROI, make the value added and acquire competitive advantage.

How to promote patient safety are the most important issues in healthcare in Taiwan. To maintain high healthcare quality is not to reach the goal in one step. To error is human. There are not the same reasons in medical error in each hospital. There is less medical adverse event in negligence by artificial to be announcement by hospitals. Now, the questions we aware are just like a nook in an iceberg. However, healthcare industry has unique characteristic about human's life. So there are maybe some "killer applications" in healthcare first. We point out the critical deployment factors by deploying RFID in healthcare. The future research tries to find out the solution to solve the questions in the paper above.

Integrating RFID with HIS will make patient more safety, and make healthcare more accuracy and efficiency, and less medical adverse event in negligence. We create an automatic healthcare mechanism to depend on five rights care quality, which include: right patient, right drugs, right dose, right time, and right route. The final vision is to create the sublime of dream by RFID assistance in the "patient safety healthcare environment".

ACKNOWLEDGEMENT

Thanks for the direct and inspire from the professor Hsin-Ginn Hwang, Cheng-Yuan Ku, Fan Wu and I-Chiu Chang in National Chung Cheng University (NCCU), and W.C Tsai in China Medical University (CMU).

And special thanks for the professor Yi-Long Huang, James Tsemeng Lin in National Tsing Hua University (NTHU); the professor Anthony Fu-Wha Han, Hsiao-Cheng D. Yu in National Chiao Tung University (NCTU); Shrane Koug Chou in National Chengchi University (NCU)

Thanks the support from dear friends in 349-1 Lab in NCCU, "Green On Digt. Inc.", Fion Chou in Oracle.

REFERENCES

- [1] Frank Kuo, Yueh-Ching Lee, Shi-Kun Ke, Chih-Ming Lee and Fan Wu, "Research issues in health industry with RFID", in Proceeding of the 3rd Asia-Pacific and Cross-Strait HL7 Conference on Healthcare Information Records, pp.148-160. Taipei, Taiwan, July 9-10, 2004.
- [2] Frank Kuo, Lee Yueh-Ching, Sang Chwei-Ching, Tsai WC, "The Research Issues of RFID in patient safety by JCAHO", e-Health Care & Care Management

- conference, pp.46~47, Kaohsiung, Taiwan, Oct2, 2004.
- [3] Frank Kuo, Nien Chu Wu, Hsiao-Cheng Yu, Cheng-Yuan Ku, "To comment healthcare industry implement RFID from HIPAA viewpoint", 2004 Technology Law Conference in Taiwan, Hsin Chu, 2004. (Accept)
- [4] Yueh-Ching Lee, Ivan Lee, Eyson Ke, Hsin-Ginn Hwang, Hsiao Ju Ling, "A feasibility study for RFID to application in supply chain management", the 3rd Conference on Services Marketing and Management, Taipei, Taiwan, June 21, 2004.
- [5] Hsin-Ginn Hwang, "Research Issues In Healthcare Information Systems," Journal of Information Management, vol. 9, pp.101~116, 2002.
- [6] To Err Is Human: Building a Safer Health System, Institute of Medicine, 2000.
- [7] Bo Tong Cheng, Communication, pp.50~59, 2004,5
- [8] Ilan Modai, Maayanit Sigler, and Rena Kurs, Clinical Computing: The Computerized Lab Alert System for Patient Management in Clinical Care, Psychiatr. Serv.; 50: 869-885, 1999.
- [9] Accenture News Release, Accenture Helps Form RFID Industry Group to Evaluate Technology's Value in Pharmaceutical Industry", Washington, US, 2004. Feb.18.
- [10] Auto-ID Center, Seize the day: The Silent Commerce Imperative, USA, 2002.
- [11] Auto-ID Center, White Paper- Towards the 5 cent Tag, 2002.
- [12] Auto-ID Center, White Paper- RFID in China, 2004.
- [13] Allied Business Intelligence Research, USA, 2003.
- [14] D.R. Reynolds, J.R. Riley, Remote-sensing, telemetric and computer-based technologies for investigating insect movement: a survey of existing and potential techniques, Computers and Electronics in Agriculture, vol. 35; pp. 271-307, 2002.
- [15] Eric Paul, Reengineering Medication Management From the Bedside Using Bar-Coding and Wireless Technology, HIMSS, 2004
- [16] EAN TAIWAN Report, Sep, 2004.
- [17] EPC/RFID International Forum 2004, Taipei, Taiwan, Oct 8, 2004.
- [18] HIMSS Board of directors, Comments on the FDA Proposed Rule- Bar Code Label for Human Drug Products and Blood, HIMSS, June 12,2003
- [19] Nien Chu Wu, RFID industry development and market trend, RF technology and industry development conference, Oct 10, 2002.
- [20] Michael, J. Spring, A supplemental distribution channel? The case of US parallel export channels", Multinational Business Review, pp. 24 -35, 1998.
- [21] Watson, R. T., L. F. Pitt, P. Berthon and G. M. Zinkhan "U-Commerce: Extending the Universe of Marketing." Journal of the Academy of Marketing Science, 30(4): 329-343, 2002.
- [22] Council For Economic Planing and Development in Executive Yuan, Challenge 2008-National Development plan, <http://www.cepd.gov.tw/2008/>
- [23] Department of Industry Technology in MOEA, technology pioneer, http://www.st-pioneer.org.tw/modules.php?name=St_News&pa=Show_News&tid=64
- [24] EAN.UCC, http://www.eantaiwan.org.tw/2003-12_3.htm
- [25] EAN Taiwan Report, September. <http://www.eantaiwan.org.tw/2003-09E01.htm>
- [26] FDA, http://www.fda.gov/cder/gmp/gmp2004/GMP_finalreport2004.htm
- [27] RFIDa, <http://www.rfida.com/data/walmartpilot.htm>
- [28] HIBCC, <http://www.hibcc.org/>
- [29] ID TechEx, <http://www.idtechex.com>
- [30] Ministry of Economic, <http://www.moea.gov.tw/>
- [31] Michelle L. Kirsche, Electronic bar codes, RFID not a matter of if, but when - radio frequency identification, Nov 17, 2003 http://www.findarticles.com/p/articles/mi_m3374/is_16_25/ai_110532978
- [32] MOEA in Department of Industry Technology, <http://doit.moea.gov.tw/news/newscontent.asp?IdxID=50&ListID=0431>
- [33] HDMA, <http://www.healthcaredistribution.org/>
- [34] Ton Yen General Hospital, <http://w3.tyh.com.tw/>
- [35] Taipei Veteran's General Hospital, <http://www.vghtpe.gov.tw/>
- [36] Taipei Medical University Hospital, <http://www.tmch.org.tw/>
- [37] Show-Chwan Healthcare System, <http://www.scmh.org.tw/>
- [38] KFAYA cancer center, <http://www.kfsyscc.org/>
- [39] Tri-Service General Hospital, <http://www.tsgh.ndmctsgh.edu.tw/>
- [40] Mao Tse-tung's memorial web, <http://www.chairmanmao.org/>

The Glance of an Observer on the War of Brands of Food Products

Carlos Alberto Gonçalves¹, Daniela Ferro de Oliveira¹,
Renata Fabiana Ferreira², Karina A. Pereira Garcia Coleta¹

¹ Business Department, Federal University of Minas Gerais, Belo Horizonte, Brazil

² Statistics Department, Federal University of Minas Gerais, Belo Horizonte, Brazil

carlos@face.ufmg.br, {daniferro, renata, karinagarcia}@cepead.face.ufmg.br

ABSTRACT

The present work proposes a model to compare means of many constructs that evaluate competitiveness of brands of the Brazilian chilled and frozen food industry. Such a model is based on a nomological network, which was built over the concepts pointed by [4] [13] [8], and [7] by the NUME – Marketing and Strategy Research Center of the Federal University of Minas Gerais - research group. Besides the validation of the new research instrument for gauging and generalization, it will be made an evaluation of brands - the ones well-positioned at the market - using the comparison of the indexes and averages of the nomological chain built for the companies, in relation to the following concepts: Tangibility (perception of packings, flavors and smells); Reliability in the Brand; Satisfaction; Loyalty; Image of the Brand; Perceived Value and Functional Conflict.

Keywords: competitiveness, food sector, nomological chain

1. INTRODUCTION

In reason of the elevated deregulation, the Brazilian food industry is highly competitive. This deregulation is a characteristic of the 'basic consuming goods', along with the short life cycle, maturity and oldness of the sector.

The food industry received, in the last years, a demand increment originated from the middle class population strata. And, with the productivity increase, the food could arrive to the final consumer with smaller costs [10].

Therefore, the emphasis doesn't just sets on the costs of the food, but it also covers the purchase services, storage, cleaning, preparation and commercialization.

In Brazil, there is a great number of companies with an assorted mix of products, which harshly dispute the customers' preference in the supermarkets' shelves.

The companies included in this research represent the leadership of the Brazilian market, they are the enterprises whose Brand names are: *Perdigão*, *Sadia* and *Seara*. The items appraised in these brands are the chilled and frozen meat products.

2. THE COMPANIES INVOLVED

Perdigão [16], founded in 1934, is a great manufacturer of poultry and pork byproducts. Along the years, the company has implemented a poultry and pork productive system, the so called vertical integration. Currently, the total number of integrated partners is of 6,810, and Perdigão has carried out a significant industrial expansion over these last 64 years. All in all the industrial complex comprises 12 meat and 2 soybean

processing units, 6 animal-feed factories, 12 incubator units and 27 company-owned poultry and pork farms.

Sadia [19] has been taking the leadership in several activities related to the food industry. It is among the largest food processing companies of Latin America, being one of Brazil's largest exporter. As a research - conducted by the English consulting firm Interbrand in 2001 - realizes, the brand Sadia was acclaimed as the most valuable company of the Brazilian food industry. Besides, according to the operational profile traced by the brokerage company Pilla Corretora de Valores Mobiliários e Câmbio, in the third quarter of 2002, Sadia was pointed as the national leader in the production and sale of frozen and chilled food products of poultry and pork meats, besides counting with the largest distribution network of frozen and chilled food products in the country.

Seara [20], founded in 1956, represents one of the largest national companies in the segment of poultry and processed meats (hams, sausages and salamis). Besides, it exports poultry cuts and pork meat. João Augusto Salles, responsible analyst for the sectors of banks and food products of the Brazilian consulting firm Lopes Filho e Associados, affirms that Seara is the largest exporter of pork of the country. It exports more pork meats than Sadia and Perdigão. Seara is a company of great stature and has its own seaport, in Santa Catarina, to export its production.

3. PROBLEM OF THE RESEARCH

The main question of this research is: *Are there meaningful differences between the averages of the brands? Does the one the brands occupy prominent position at all the constructs of clients' perceptions?*

4. OBJECTIVES OF THE RESEARCH

With this research, the goal is to build a tool that could measure differences between competitive brands and prepare a new measure tool to the *commodity behavior sector*.

5. CONSTRUCTS OF THE RESEARCH

In the present research there are constructs of different models. From the Model SERVQUAL [2], the constructs Tangible Aspects and Reliability were used. From the ACSIndex - American Consumer Satisfaction Index, proposed by [4], the constructs Loyalty, Satisfaction and Perceived Value were taken. The construct Image was retrieved of the conceptualization of [1], while the Functional Conflict came from [13] Relationship Model.

5.1. Tangible Aspects and Reliability: The SERVQUAL Model

The first studies about quality of services were of authorship of [14] and had as objective to search for an integrative model in that area. Therefore, the managers and customers of four North American companies were questioned on the fundamental attributes of the service quality, being also brought up the existence of divergences between the two opinions. The following companies participated in the research: retail banks, credit card administrators, property brokers, and repair and maintenance of goods firms.

In a second phase of the research, [15] focused their studies in the measurement of Gap 5 (gap between the perceived and the expected service), appearing the famous equation:

$$Q = P - E \quad (1)$$

In other words, Quality = Perception - Expectations. Initially 97 items, referring to the external dimensions of the service quality, were generated. Later, the scale was refined and it came to a scale of 22 items. The grid of items contained in the research instrument reflected the following dimensions: Tangible Aspects, Reliability, Promptness, Guaranty and Empathy. Such dimensions are:

- Tangible Aspects: physical facilities, equipments and appearance of the company's personnel. When the consumer enters in contact with the atmosphere of a supplying company.
- Reliability: capacity and ability to implement the promised service in a safe and reliable way.
- Promptness: good will to aid the consumer and to provide ready attendance - to solve problems on time.

- Guaranty: employees' knowledge and courtesy and their ability to inspire credibility and trust - to assert that the service is safe and guaranteed.

- Empathy: individualized consideration and attention that the company renders to its consumers - fine-tuning of sympathy and understanding between supplier and customer.

In this study, only the Tangible and the Reliability Aspects were explored.

5.2. Loyalty, Satisfaction and Perceived Value: ACSIndex

The Model of the ACSIndex - American Consumer Satisfaction Index - is proposed by [5]. That model intends to offer a base of uniform and comparable measurement for the customer's global satisfaction, besides pointing relationships of such construct with its main antecedents and consequents. Implicit in the model is the recognition that the customer's global satisfaction cannot be directly measured, being a latent variable requesting multiples indicators in its measurement. The most immediate and tangible result of the operationalization of that model is a score of the latent variable of the customer's global satisfaction, in terms, generic enough, for a comparison among supplying organizations, branches of activities, sectors and nations.

5.2.1. Loyalty

The final relationship of the model is between the customer's complaints and their loyalty. The direction and the indication of that relationship depends on the service systems rendered to the customer and on the solution - by the supplier - of the clients complaints [4]. When the relationship is positive, the implication is that the supplier succeeded in transforming a customer that complains into a loyal customer. When the relationship is negative, the supplier worked with the situation in such a way that the negative situation became even worse, contributing to the loss of the customer's loyalty.

5.2.2. Satisfaction

The customer's global satisfaction, as the central construct of the model is placed inside of a relationship chain that goes from its antecedents (expectations, quality and value perceived by the customer) to its consequences (complaints and the customer's loyalty). Of special interest in the model, beyond its own global satisfaction, is the explanation of the customer's loyalty, as very probable indicator of profitability [17]. With that structure, the model allows the ACSIndex to be tested under the nomological point of view. Nomological validity is the degree in that a construct behaves as predicted inside of a system of related constructs, the so-called nomological network [3].

5.2.3. Perceived Value

A second determinant of the customer's global satisfaction is the perceived value. This is the product perceived level of quality of the product in relation to its price. The factor price is incorporated to the perceived value, reinforcing the comparability of results among suppliers, branches of activities and sectors. It is supposed to be a positive association between the product's perceived value and the customer's global satisfaction.

5.3. Image: Lalande / Barich and Kotler

The author [11] defines image as being the mental repetition, usually weakened, of a sensation (or more exactly of a perception) previously experienced. Image can then, be considered, as being a certain way of appropriation of the reality for a certain subject, in other words, as perception phenomenon. Perception can be understood, according to [11], as an act in which the individual - organizing its present sensations, interpreting them, and complementing them with images and memories - opposes to an object that he/she spontaneously considers as different from itself - real and unknown.

Through the perception happens an internalization process - by the individual - of the received stimuli, that will make possible the formation of the image, through which he/she will recognize such incentives. The perception phenomenon is, according to [11], the identification of the reality; happening after the sensation phenomenon, when the individual will learn - through a selective process - some values present stimuli received and will aggregate to it other subjective values and internal objectives.

In that way, image can be considered as being the subjective vision of the objective reality. The reception of the emitted message will be formed, starting from the process of interpretation of that message, being then, modified, and acquiring its own version in the individual's conscience [11].

The image that the individual has of the real (symbolic and different vision from the reality) unchains an attitude of that individual towards the object - therefore, the image has then, the power to influence the individual's behavior [18].

The first image focuses linked to marketing appeared in the 1950s. The authors [6] verified that the consumers not only valued the physical, tangible aspect of the products they buy, but also the symbolic meanings attached to the brand of those products.

6. MODEL OF THE RESEARCH

Once the constructs have already been explicated, the adapted model is in the FIG.1.

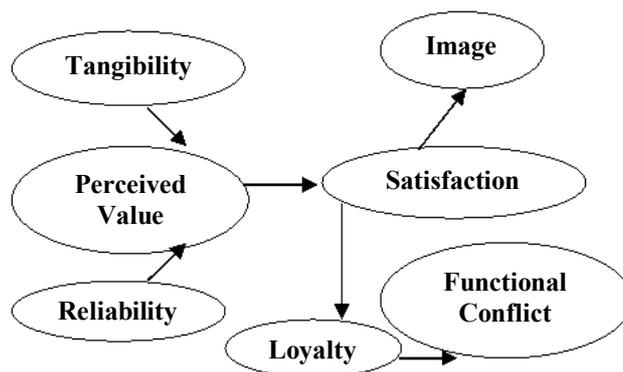


Figure 1 - Adapted model

7. ANALYSIS RESULTS

7.1. Factor Analysis

In order to analyze interrelationships among a large number of variables, Factor Analysis was chosen as the statistical approach in this work. According to [9], the aim is to explain these variables in terms of their common dimensions, called factors. Hence, factors are dimensions that try to explicit the existent correlations between a group of variables. This process enables the researcher to lose the minimum of information.

A multivariate method included in the group of Factor Analysis is the Principal Component Analysis, which derives factors that contain small proportions of unique variance [9].

This method is recommended in case the objective is to determine the minimum number of factors (principal components), that answer for the maximum variance on the data for future multivariate analysis uses [12].

It is worth to remember that in TAB. 1 and 2, only the factor loadings for each question higher than 0.5 were kept in the table, showing a good correlation with each factor [9].

TAB.1 shows the Rotated Component Matrix. It shows how many factors were built by the collected data. Six factors were found, and this is a good result, once there are 7 constructs in this research. Only the loadings over 0,5 were maintained in the matrix, according to [9].

From the TAB.1, it's able to see some well dimensioned factor and others not so much. The first factor grouped several indicators from three different constructs (Tangibility, Reliability, and Loyalty). The second factor is related again with Tangibility and Reliability, both from the SERVQUAL model of [15]. Functional Conflict is totally explained by the third factor. The fourth factor is fragmented between Image and Satisfaction, with the fifth is clearly Perceived Value. The sixth is more related to Tangibility.

In order to reach a better result, a seven-factor rotated component matrix was forced in future factor analysis, as TAB. 2 displays.

Table 1 - Rotated Component Matrix

	Factors					
	1	2	3	4	5	6
The products packages are practical and easily handled						,556
The information in the packages correspond to what the products represent						,535
I consider the products of this company flavorful		,580				
The ready foods to eat have a pleasant smell		,748				
The color of the products of this company is pleasant		,783				
I believe in the quality of the products from this company		,607				
The products are safely packaged	,612					
The company keep their promises in its products	,586					
I recommend this company's products to my family and friends	,741					
When I come across new products of this company, I do not hesitate in buying and consuming them				,533		
I feel comfortable in recommending the products of this company to other people	,734					
I defend the products of this company when somebody makes negative commentaries about them						
If the press releases good news about this company, I am immediately inclined to believe in it	,514					
The logo of the company means a serious presentation to me, and of good reputation				,512		
When I go shopping, I search in order to identify the products of this company on the shelf						
For me, this company has a positive image				,606		
I am satisfied with the existence of a company like this supplying food				,732		

	Factors					
	1	2	3	4	5	6
If there aren't any offers of products from this company, I will not hesitate in buying products from another company						-,750
The products of this company satisfy my expectations				,722		
I believe the price charged for the products of this company is fair, considering the quality offered					,844	
The price I pay for the products from this company is within my expectations					,882	
I believe that, if I detect some problem in a product, the company will solve it promptly				,810		
I believe that my suggestions to improve the products will be Heard by the company				,823		
I believe that the company, in case I complain, will pay attention to me				,918		

Table 2 - Rotated Component Matrix – Forcing 7 factors

	Factors						
	1	2	3	4	5	6	7
The products packages are practical and easily handled							,555
The information in the packages correspond to what the products represent							,575
The products are safely packaged	,661						
I consider the products of this company flavorful	,539						
The ready foods to eat have a pleasant smell				,820			
The color of the products of this company is pleasant				,750			
I believe in the quality of the products from this company	,692						

	Factors						
	1	2	3	4	5	6	7
The company keep their promises in its products	,687						
I recommend this company's products to my family and friends	,606				,521		
When I come across new products of this company, I do not hesitate in buying and consuming them							
I feel comfortable in recommending the products of this company to other people	,688						
I defend the products of this company when somebody makes negative commentaries about them					,712		
If the press releases good news about this company, I am immediately inclined to believe in it					,655		
The logo of the company means a serious presentation to me, and of good reputation			,512				
When I go shopping, I search in order to identify the products of this company on the shelf							
For me, this company has a positive image			,614				
I am satisfied with the existence of a company like this supplying food			,766				
If there aren't any offers of products from this company, I will not hesitate in buying products from another company							-,742
The products of this company satisfy my expectations			,723				

	Factors						
	1	2	3	4	5	6	7
I believe the price charged for the products of this company is fair, considering the quality offered						,856	
The price I pay for the products from this company is within my expectations						,893	
I believe that, if I detect some problem in a product, the company will solve it promptly		,812					
I believe that my suggestions to improve the products will be Heard by the company		,819					
I believe that the company, in case I complain, will pay attention to me		,917					

7.2 Means Comparison

For a means comparison, it was used the Analysis of Variance (ANOVA), a method to test the equality between means from three or more groups [21]. The hypothesis tested in ANOVA is related to the means between the different groups. If the calculated P Value is less than 0,05, one concludes, with 95% reliability, that at least one of the groups has the different mean compared to the others.

The ANOVA only verifies if there are significant differences between the groups; so, to point out which of them are different, it was used the Duncan's Multiple-Range Test, that compares all the pairs of involved means in a study of Analysis of Variance [22].

TAB. 3 shows that one company holds the best position in the ranking in relation to almost all the constructs, and three are tied up in the constructs Perceived Value and Functional Conflict. This shows that to invert the order it is necessary a significant investment from the company in the second position and even more from the third position, so that the company can enter the battle to be the first option of the customer. An option that the first position still has is to also be hegemonic in the two constructs where they are tied up, once the company looks for investments in this direction. Through a deep look on the TAB.3, it is easy to see that the leading company in the food sector (Sadia) has an average of perception, in almost all constructs, higher than the other companies. There was no significant difference

between Perceived Value and Functional Conflict among the three companies.

Table 3 – ANOVA

	Sadia	Perdigão	Seara	ANOVA (P Value)	Duncan's Method
TANG	24,8231	26,8719	28,0839	0,000	Sadia > Perdigão > Seara
RELIAB	18,3438	17,6283	15,8081	0,000	Sadia > Perdigão > Seara
LOYAL	12,4431	11,8700	10,2425	0,000	Sadia = Perdigão > Seara
IMAG	14,5600	13,7001	12,1394	0,000	Sadia > Perdigão > Seara
SATIS	12,9963	12,4318	11,4378	0,000	Sadia > Perdigão > Seara
PVAL	6,6775	6,6570	6,4005	0,676	There isn't significant difference
FUNCC	13,0395	12,9455	12,1006	0,063	There isn't significant difference

Key: Tang – Tangibility; Reliab - Reliability; Loyal – Loyalty; Imag – Image; Satis – Satisfaction; Pval – Perceived Value; Funcc – Functional Conflict.

7.3 Multiple Regression Analysis

Multiple Regression Analysis is a statistical tool used to measure the relationship between one dependent variable and several independent variables, called predictors [9]. The objective is to analyze the possible strong and weak relations between constructs. The regression equation's adjust quality is verified through R Square, which is the model for coefficient of determination. The method used in this analysis was Stepwise, that takes out the non-significant variables from the model.

TAB. 4 presents the regression equation with standardized coefficients, having the construct Image as the dependent variable and Tangibility, Reliability, Satisfaction and Quality as the predicting variables. In this context, it is important to observe the emphasis or major weight attributed to the predictors by the companies, in order to build the Image organization in the competitive market.

Through the results in the TAB. 4, we can observe that Seara suggests that builds its Image through its clients' Satisfaction. Besides, it emphasizes the Satisfaction perception with more emphasis than Perdigão and Sadia in building the Image construct, unlike the other companies, that give more importance to Tangibility than Seara does.

Table 4 - Regression analysis for the Image of the companies - Standardized coefficients

Perdigão	0,241 (0,000)	0,325 (0,000)	0,118 (0,044)	0,987	0,000
Sadia	0,222 (0,000)	0,318 (0,000)	0,188 (0,000)	0,993	0,000
Seara	0,125 (0,013)	0,317 (0,000)	0,354 (0,000)	0,977	0,000

TAB. 5 presents the regression equation with standardized coefficients, having the construct Loyalty as the dependent variable and Reliability, Perceived Value and Quality as the predicting variables. In this context, it is important to observe the emphasis or major weight attributed to the predictors by the companies, in order to build the Loyalty organization in the competitive market.

Table 5 - Regression analysis for the Loyalty of the companies - Standardized coefficients

Perdigão	0,599 (0,000)	0,201 (0,013)	0,965	0,000
Sadia	0,601 (0,000)	0,216 (0,011)	0,964	0,000
Seara	0,649 (0,000)		0,952	0,000

As we can see through the results showing in the table 5, the Loyalty to Perdigão and Sadia are based not only on Reliability, but also on Perceived Value, while Loyalty for Seara is directly linked only to Reliability.

Clients' Satisfaction for Sadia is more related to Reliability than the other companies, once Satisfaction for the others is also related to Tangibility.

Functional Conflict for Perdigão is more related to Tangibility, and for Seara is linked to Perceived Value. Only Sadia links Loyalty to Functional Conflict.

Table 6 - Regression analysis for the Satisfaction of the companies

Perdigão	0,497 (0,000)	0,132 (0,011)	0,969	0,000
Sadia	0,703 (0,000)		0,971	0,000
Seara	0,470 (0,000)	0,159 (0,009)	0,959	0,000

Table 7 - Regression analysis for the Functional Conflict of the companies

Perdigão	0,425 (0,000)	0,216 (0,028)		0,956	0,000
Sadia	0,324 (0,000)	0,196 (0,047)	0,207 (0,020)	0,957	0,000
Seara	0,322 (0,000)	0,624 (0,000)		0,944	0,000

7.4. Structural Equations Modeling

SEM (*Structural Equation Modeling*) is a tool used in both academic and managerial research. According to [9], it is a multivariate technique which combines aspects of multiple regression analysis – examining dependence relations – and factorial analysis, in order to estimate several dependence relations interrelated simultaneously.

In this paper, the technique was used to verify – grouping the researched companies – the influence level of each construct in the other ones, and it also investigates stronger dependency relations from the proposed model.

FIG.2 presents the model with non-standardized coefficients which correspond to regression weights in multiple regression and they are expressed in terms of construct scale, variance in this case [9]. As scale varies from one construct to another, the comparison among coefficients becomes more difficult than using standardized coefficients (FIG. 3).

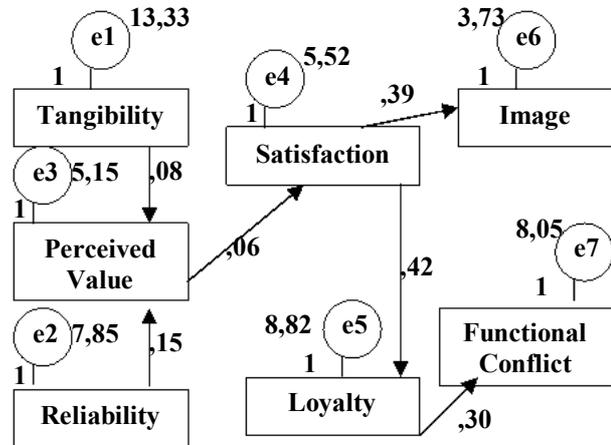


Figure 2 – Model with non-standardized coefficients

Model goodness of fit was unfavorable because of the results of Chi-Square Test and P Value lower than 5% (considering as a desirable value of significance level: Chi-Square = 665,197; Degrees of Freedom = 15; P Value = 0,000). Thus, the proposed theoretical model – in theoretical nomological chain – is considered not to be adjusted with empirical measurements.

In a general way, correlation coefficients presented are low and indicate a weak adjustment of theoretical linear model to the empirical data behavior. This work understands this as a limitation that should be investigated in a deeper way including new attempts of data mining in order to rotate the model with transformed variables. In spite of this limitation, we can see that for the values presented by the arrows (FIG.2) the biggest regression coefficient is the one that goes from Satisfaction to Loyalty, it means that Satisfaction explains 42% of Loyalty. Beta coefficient non-standardized from Satisfaction to Image is also high, thus Image is explained by 39% of the first construct. From those observations, it is possible to infer, with some limitations, that clients' Satisfaction for all the three companies implies in a higher Loyalty and it also contributes to construct a better company Image.

FIG.3 presents standardized coefficients where all of them have equal variances and they are useful to determine relative importance, but they are specific to each sample, so they are not used to comparison among samples (as non-standardized coefficients).

With standardized coefficients it is possible to verify that Satisfaction explains 42% of Image constructed by clients of researched companies, important managerial information. The number above "Image" box (.18) indicates that antecedent constructs together, in a Multiple Linear Regression, explain 18% of Image.

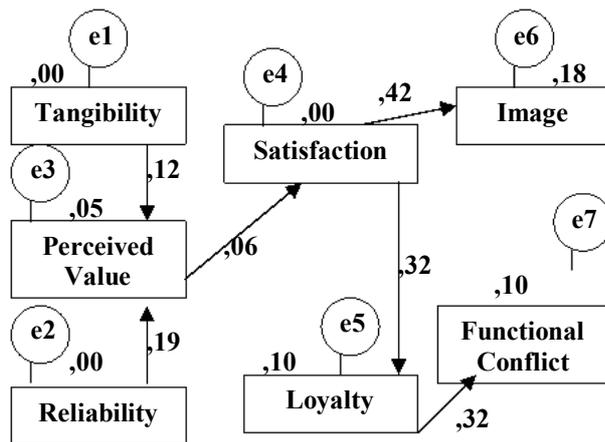


Figure 3 - Model with standardized coefficients

8. CONCLUSIONS

Analyzing all the results, it's clear that the perception of the respondents is equal to reality: Sadia continues to be the leader in the nourishing sector. This consideration is very important, once people aren't always able to differ objectively what is real in the market.

As we can see, Sadia detaches in the preference of consumers – the reasons are that Sadia has been applying resources in product development and quality in the Brazil market. Nowadays, it's becoming very difficult for the other competitors to obtain the leading position. The competitors keep launching products and innovating in a competitive market, but time has showed a stable position of Sadia in relation to the preference of consumers.

In Regression Analysis, it was confirmed that Sadia continues to be the leader in this food sector, and another information was observed (TAB. 4). This company gives more importance to Reliability to build its Image, although Perdigão, the second in ranking, has the higher coefficient in the analysis.

It was also observed that Loyalty really depends on a good level of Reliability, for all the three companies, important information to management decisions.

Structural Equations Modeling presented strong relations among constructs Satisfaction, Image and Loyalty. Thus, a satisfied client becomes loyal and builds a good company image.

In future studies, an experiment can try to measure possible unbalancing of this positioning order (commodities unbalancing).

REFERENCES

[1] BARICH, Howard & KOTLER, Philip. *A framework for marketing image management*. *Sloan Management review*. V.32,n.2, p. 94-104. winter, 1991.

- [2] BERRY, Leonard L., PARASURAMAN, A. *Serviços de Marketing: competindo através da qualidade*. São Paulo: Maltese-Norma, 1992, 238p.
- [3] CRONBACH, Lee J. & MEEHL, Paul E. Construct Validity in Psychological Tests. *Psychological Bulletin*. vol. 52, n. 4, p. 281-302, 1955.
- [4] FORNELL, Claes. A National Customer Satisfaction Barometer: The Swedish Experience. *Journal of Marketing*, vol. 56, p. 6-21, Jan. 1992.
- [5] FORNELL, C.; JOHNSON, M. D.; ANDERSON, E. W.; CHA, J. & BRYANT, E. The American Customer Satisfaction Index: nature, purpose, and findings. *Journal of Marketing*, v.60, n. , p.7-18, Oct 1996.
- [6] GARDNER, burling. B.E. & LEVY, Sidney j. *The product and the brand*. *Harvard Business Review*. V. 33, p. 33-9, marc/apr,1995.
- [7] GONÇALVES, C. A., GARCIA, K. A. P., OLIVEIRA, D. F. *Avaliação das Relações entre os Construtos Ligados à Satisfação Global do Cliente, ao Relacionamento e à Imagem – Uma Verificação B2C dos Usuários de Telefonía Móvel*. Relatório de Pesquisa FAPEMIG: Belo Horizonte, 2002.
- [8] GOSLING, Marlusa. *Estratégias de Relacionamento no Setor Bancário Brasileiro: um Estudo Empírico*. Dissertação de Mestrado. Belo Horizonte: CEPEAD, UFMG, 2001.
- [9] HAIR, Joseph F. et al. *Multivariate Data Analyses*. New Jersey: Prentice Hall, 1998.
- [10] INVESTSHOP. Available from <http://www.investshop.com.br>. Cited: 10 Jan. 2003.
- [11] LALANDE, André. *Vocabulário Técnico y Crítico de la Filosofia*. Buenos Aires: El Ateneo, 1966.
- [12] MALHOTRA, Naresh K. *Pesquisa de Marketing: uma orientação aplicada*. Porto Alegre: Bookman, 2001.
- [13] MORGAN, Robert M.; HUNT, Shelby D. The commitment-trust theory of relationship marketing. *Journal of Marketing*, v. 58, p. 20-38, Jul., 1994.
- [14] PARASURAMAN, A , ZEITHAML, V., BERRY, L. A Conceptual Model of Service Quality and Its Implications for Future Research. *Journal of Marketing*, vol. 49 (Fall 1985), 41-50.
- [15] PARASURAMAN, A , ZEITHAML, V., BERRY, L. SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, vol. 64, number 1, 12-40, New York University, Spring 1988.
- [16] PERDIGÃO. Available from <http://www.perdigao.com.br>. Cited: 10 Jan. 2003.
- [17] REICHHELD, Frederick F. & SASSER, W. Earl. Zero Defections: Quality Comes to Services. *Harvard Business Review*. vol. 68. p. 105-11. Sep.-Oct. 1990.
- [18] REIS, Maria do Carmo S. *Imagem Corporativa: Produção e Consumo*. Belo Horizonte: Faculdade de Ciências Econômicas da UFMG, 1991. (Dissertação, Mestrado em Administração).
- [19] SADIA. Available from <http://www.sadia.com.br>. Cited: 10 Jan. 2003.

[20] SEARA. Available from <http://www.seara.com.br>. Cited: 10 Jan. 2003.

[21] TRIOLA, Mario F. Introdução à Estatística. Rio de Janeiro: LTC, 1999.

[22] WERKEMA, Maria C. C., AGUIAR, S. Planejamento e Análise de Experimentos: como identificar as principais

variáveis influentes em um processo. Belo Horizonte: Fundação Cristiano Ottoni, Escola de Engenharia da UFMG, 1998, vol.8.

Theoretical Framework of Empirical Study on Consumer Behavior in Web-Based Commerce

Yuming Zhu, Peng Guo

Management School, Northwestern Polytechnical University, Xi'an 710072, China
{zym1886, guopeng}@nwpu.edu.cn

ABSTRACT

Web-based commerce has brought great challenges and opportunities to companies. To seize the opportunities, the company must work out well-conceived web-based marketing strategy, which, to great extent, relies on the good understanding of consumer behavior in web-based commerce. Based on general theoretical framework of study on consumer behavior and current achievement in this area both oversea and home, the theoretical framework of empirical study on consumer behavior in web-based commerce is put forth. In this theoretical framework, two dependent variables is determined as unplanned purchase and customers' intention to return, which are affected by two attitudinal variables, i.e. perceived control and shopping enjoyment. The two attitudinal variables are also affected by individual factors and web store factors, which also contain certain sub-factors. At the end of this paper, some suggestions on further study in this field are brought forward as well.

Keywords: theoretical framework, empirical study, consumer behavior, web-based commerce

1. INTRODUCTION

Along with high speed development of net technology and globalization, the portion of on-line sales in retail sector will become bigger and bigger. According to the information from U.S. Commerce Ministry at Nov. 21, 2003, the total on-line sales of U.S. was increased by 27% up to \$13.3 billion during the third quarter. In Europe, the on-line sales of 2003 are up to 70 billion Euros. In China, according to the survey of iResearch, the total on-line sales of 2003 were RMB0.79 billion. For 2004, the sales will exceed RMB1.6 billion^[1]. Therefore, the web-base consumption has been bringing revolutionary changes and impacts on traditional ways of consumption. In business-to-consumer electronic commerce there have been two transformations: (1) the transformation of the consumer into a computer user, and (2) the transformation of the physical store into a phenomenon that is information technology intensive—in other words, a Web site. The new double role of consumer/computer user resulting from the first transformation is inherently complex and still not well understood. The second transformation concerns store technology. In the physical commercial world, the information technology used in a store is most often invisible to the consumer, whereas in e-commerce it has been moved to the foreground. To the consumer, the Web site, an information system, is a full representation of the store. Therefore, it is necessary to look at the interaction between consumer and store through the filter of user-technology interaction. Given these transformations, researchers need to combine the research streams of information systems, psychology, and marketing to study and understand on-line consumer behavior. The empirical study, which is widely used by western researchers, is just the required combination of studying. However, due to different environment, some results achieved through empirical study by western

researchers cannot be directly and simply applied in China.

2. ACHIEVEMENT ON CONSUMER BEHAVIOR ON WEB-BASED COMMERCE

At present, most of successful B2C companies are Western companies, such as AMAZON, DELL etc. There are two causes for this: (1) well-established credit system of paying, and (2) well-combined research and practice. By empirical study, the western researcher found that perceived control and shopping enjoyment can increase the intention of new Web customers to return, but seemingly does not influence repeat customers to return. They also found that a Web store that utilizes value-added search mechanisms and presents a positively challenging experience can increase customers' shopping enjoyment. Further, the more often customers return to a Web store, the more their shopping enjoyment is determined by their product involvement. Customers with low need specificity (i.e., who do not know what they are looking for) are more likely to use value-added search mechanisms. Finally, neither perceived control nor shopping enjoyment has any significant impact on unplanned purchases^[3].

In China, the researchers studied consumer behavior on web-base commerce from qualitative aspect and quantitative aspect respectively.

From qualitative aspect, the researchers found the following changes which have taken place in consumers' psychology: (1) personalized consumption prevailed; (2) initiative consumption increased; (3) stability of consumption psychology decreased, rapid switching speed; (4) chasing both convenience and enjoyment of consumption; (5) price still be one of most important factors. However, owing to characteristics of web-based

consumption, there must be certain psychological needs cannot be satisfied. For example, due to partial replacement of personal intercommunication, the social psychological needs can not be complied with to the full. In addition, the on-line shop cannot give consumer the feeling of being noticed and respected. The consumers cannot show their social status, achievement and dominance. Furthermore, less flexibility of pricing lead to dissatisfaction of consumers who like bargaining on the scene^[4].

From quantitative aspect, the researchers also make certain achievement such as *web-based consumption: theoretical model and behavior analysis*, which is a book to sum up the findings of researchers of Harbin Institute of Technology. By using this mathematic model, the researchers found that the process of on-line consumption is a combined process of production and consumption, and a complicated process of ensuing of utility increase and decrease. Chinese researchers have also set up another mathematic model called TPB-based (Theory of Planned Behavior) and E-commerce oriented theoretical model for consumer behavior study. This model has certain constructive effects on E-commerce. But, it still has following pitfalls: (1) some variables, such as social hierarchy and culture difference, are hardly quantified; (2) the combination of qualitative simulation and quantitative one is hardly fulfilled^[5].

In china, the empirical study is seldom used in this area. The on-line investigations mainly focus on the areas such as types of users, purpose of surfing internet and comments on certain problems etc.

3. THEORETICAL FRAMEWORK OF EMPIRICAL STUDY ON CONSUMER BEHAVIOR IN WEB-BASED COMMERCE

There are two factors which have major impact on consumers' behavior. One is individual factors such as needs, awareness of brand identity and attitude toward brand selection. In addition, demographic factors, life style and personality have impacts on brand selection either. The other is environmental factors such as culture, sub-culture and direct contact group etc. Marketing organization also form an integral part of environmental factors. Based on these two factors, the general theoretical framework of study on consumer behavior can be set up, which is depicted in Figure 1.

Environmental and individual characteristics are known to influence emotional responses that, in turn, affect behavioral responses. In the area of environmental psychology, three key emotional variables that may determine behavior are pleasure, dominance, and arousal. These variables are also important determinants of consumer behavior off-and on-line. In a consumer context, pleasure can be defined as shopping enjoyment, and dominance as the equivalent of the perceived control felt by customers. Arousal deals with more physical sensations, such as being jittery, sluggish, or relaxed. In this study, it was not possible to directly measure physical responses to the on-line shopping experience, and instead the research concentrated on the first two variables.

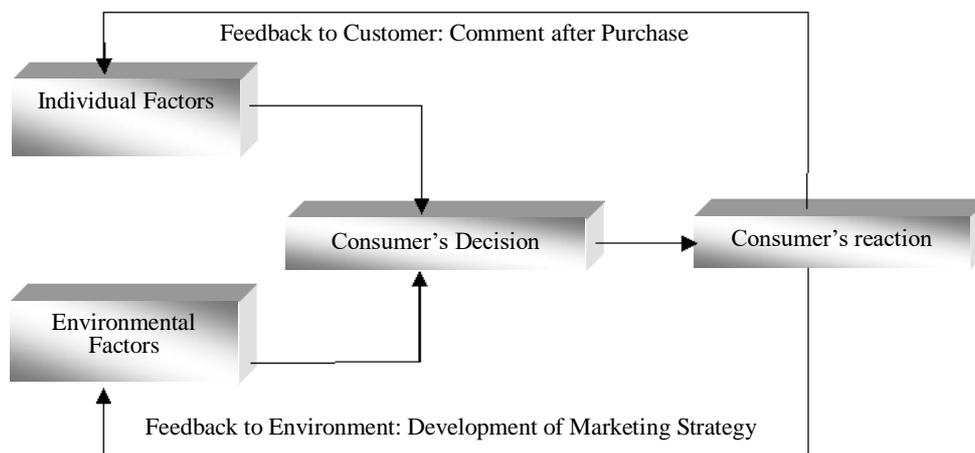


Figure 1 General Theoretical Framework of Study on Consumer Behavior

The theoretical framework of empirical study on consumer behavior in web-based commerce, which is depicted in Figure 2, includes shopping enjoyment and perceived control as attitudinal variables that directly affect on-line consumer behavior. In accordance with

both environmental psychology and traditional consumer behavior, it also examines the direct or moderating impact of individual and Web store (environmental) factors on consumer attitudes and behavior.

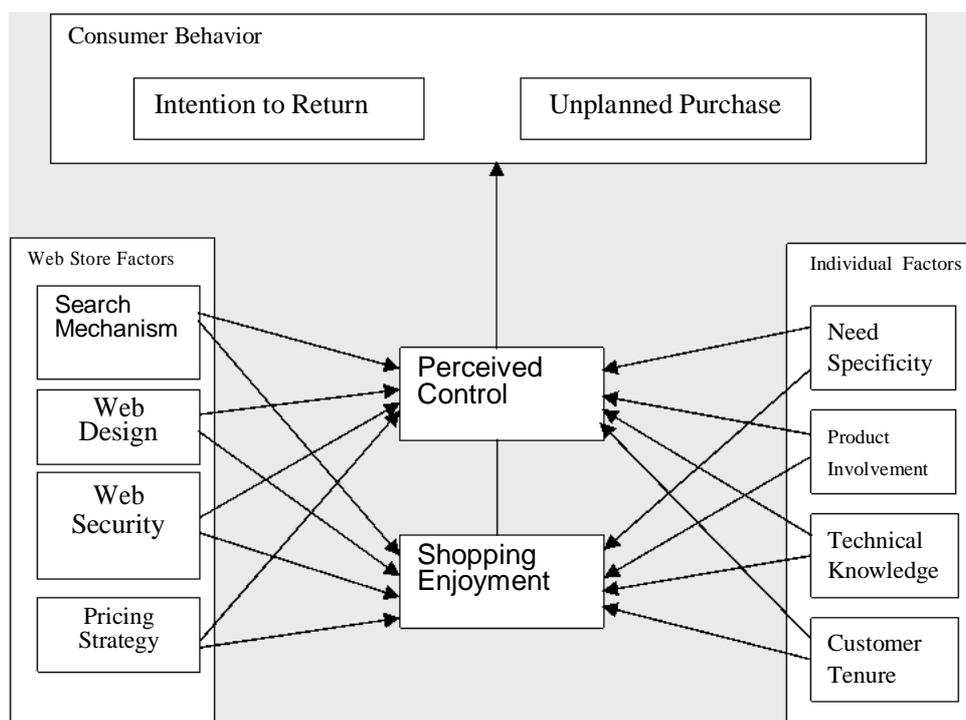


Figure 1. Theoretical Framework for Web-based Consumer Attitudes and Behavior

3.1 Dependent Variables

In this theoretical framework, customers' behavior is classified into two types (dependent variables). One is unplanned purchase. In some situations on the Web, conditions that enable a consumer to save time could lead to more unplanned purchases. The ability to search more products, services, and information may lead to longer purchasing decisions with a negative impact on unplanned purchases. Money availability is as important on-line as off-line. The other is customers' intention to return (repeat purchase). Repeat purchase for Web-based companies can decrease because consumers do not have to spend substantial amounts of money or time for transportation from store to store. They can buy from any company available on-line while sitting in front of their computer. Also, discretionary loyalty is no longer necessary. The center for one-stop shopping is now the customer's computer, where switching from one store to another is easy, not requiring significant physical effort, though it may require some learning of the new Web site. For consumers who dislike shopping, consumer loyalty on the Web may not decrease significantly, since even the small switching costs on the Web can prove too much for them. All of these factors suggest that the design of satisfactory experiences will be crucial to store loyalty.

3.2 Attitudinal Variables

Generally, the above-mentioned two kinds of consumers' behavior are affected by two attitudinal variables, i.e. perceived control and shopping enjoyment. Web stores have responded to the desire for customer control and convenience with systems that enable

consumers to easily find what they need, learn more about it, and quickly make a purchase decision. Internal search engines, hierarchical classifications of company products, and intelligent agents are provided by many Web stores to support users' need for control in accessing information. Sites offer quick, automated purchasing through "one-click" buying and the use of shopping carts. All these site features enable Web customers to enjoy higher levels of control and convenience. Shopping on-line is a different experience from shopping in a physical retail store. One major point of difference deals with "store atmosphere". This term describes the physical aspects of a store, such as colors, music type, music volume and tempo, and layout of products. Store atmospherics have a direct effect on customer mood and behavior. The atmosphere of Web-based businesses is at best limited to the confines of a computer monitor that usually displays only two-dimensional pictures and text. Even with the addition of three-dimensional images and musical accompaniment, Web stores today cannot fully simulate the ambiance of a physical store. In the competition between click and mortar, the system design of the e-tailing experience must compensate for the loss of traditional in-store atmosphere.

3.3 Individual and Web Store Factors

Obviously, the attitudinal variables are affected by individual and web store factors. Individual factors include the following sub-factors: (1) **Need Specificity**: in other words, how well individual consumers know what they want when they visit a store; (2) **Product Involvement**: the concept comprises (a) one's motivational state (i.e., arousal, interest, drive) toward

an object where (b) that motivational state is activated by the relevance or importance of the object^[6]; (3) Technical Knowledge: how well individual consumers know what kind of skills they need to surf internet smoothly; (4) Customer Tenure: how many times a customer has been to a Web store in the past. Web store factors include the following sub-factors: (1) Search Mechanism: i.e. site search features, such as internal search engines, hierarchical classifications, and intelligent agents etc.; (2) Web Design: attractiveness of web site; (3) Web Security: the safety of flow of information, cash and goods etc.; (4) Pricing Strategies, such as discount pricing strategy, reputation pricing strategy and personalized pricing strategy etc.

4. CONCLUSIONS AND DIRECTION OF FURTHER RESEARCH

Based on general theoretical framework of study on consumer behavior and current achievement in this area both oversea and home, the theoretical framework of empirical study on consumer behavior in web-based commerce is put forth. In this theoretical framework, two dependent variables is determined as unplanned purchase and customers' intention to return, which are affected by two attitudinal variables, i.e. perceived control and shopping enjoyment. The two attitudinal variables are also affected by individual factors and web store factors, which also contain certain sub-factors.

Through this theoretical framework, the factors that influence customers' on-line behavior can be determined. However, the influence rate of each and every factor still remains unclear. To solve this problem, the empirical study based on this theoretical framework should be adopted. One of the best methods of empirical study is by on-line questionnaire. Through questionnaire and with the help of statistical software, the significance of each factor can be measured by regression analysis.

Modern technology equips consumers by giving them accessibility to unprecedented information and communications. As customers' switching costs goes down, the challenges of attracting, retaining, and realizing profits from customer interactions increase. Well-conceived studies that integrate marketing perspectives on consumer behavior and information systems research on users' behaviors will be crucial to

understanding consumers' on-line behaviors. Indeed, each of the major variable categories considered in this theoretical framework, i.e. individual consumer factors, Web site design factors, consumer/user attitudes and decision processes, and behavior, presents opportunities for marketing and information systems researchers. Further research on these topics from a multidisciplinary perspective will be critical to the effective design of systems that strategically influence customer behavior.

REFERENCES

- [1] Chunru Li, "2003: E-commerce emerge again", *Information and computer*, pp35-40, 2004.2 (in Chinese).
- [2] Henry Assael, *consumer behavior and marketing action*(6th edition), China Machine Press, 2000
- [3] Marios Koufaris, Ajit Kambil, and Priscilla Ann LaBarbera. "Consumer Behavior in Web-Based Commerce: An Empirical Study", *International Journal of Electronic Commerce*, Vol. 6, No. 2, pp121-145, Winter 2001-2002.
- [4] Jing Zhuang, "Internet marketing from psychological perspective", *Academic Journal of liaoning Teacher School*, pp24-26, 2001.5 (in Chinese).
- [5] Zhicheng Li, Meilian Liu, "Study on consumers' behavior in e-commerce environment", *China Management Science*, pp56-59, 2002.6 (in Chinese).
- [6] Mittal, B. "A theoretical analysis of two recent measures of involvement. In T.K. Srull (ed.)", *Advances in Consumer Research*, vol. 16. Provo, UT: Association for Consumer Research, pp697-702, 1989.
- [7] Yuemei Jia, "Characteristics of consumer behavior in internet era and marketing strategy", *Modern Finance*, pp23-26, 2001.8 (in Chinese).
- [8] McQuivey, J.L.; Delhagen, K.; Levin, K.; and Kadison, M.L. "Retail's growth spiral", *Forrester Report*, 1, 8 (1998).
- [9] Li Liu, "Characteristics of consumer behavior in IT environment", *Journal of Information*, Vol. 21, No.2, pp45-47, 2002 (in Chinese).
- [10] Huaizhu Li, *Research Methodology for Management*, Xian Jiaotong University Press, 2004

Willingness to Pay for Digital Services: Challenges for Future Research

Herbjorn Nysveen¹, Per E. Pedersen²

¹Norwegian School of Economics and Business Administration

²Per E. Pedersen, Agder University College

ABSTRACT

In this article we discuss three issues that should be considered important in future studies of willingness to pay (WTP) for services in digital networks. First, existing studies of WTP for digital services have several weaknesses when it comes to context- and incentive realism and we suggest alternative methods to overcome these weaknesses. Second, a review of existing studies of WTP for digital services also reveal that existing studies are narrowly focused on measures of willingness to pay. Thus, a nomological network of WTP is suggested to get a broader theoretical framework for purposes of validation of pure willingness to pay measures. Finally, existing studies on WTP for digital services has focused single-channel services. The large number of cross-channel and multi-channel services offered in digital networks, makes us point to the importance of WTP studies for cross-channel and multi-channel digital services in the future.

1 INTRODUCTION

Consumers' spending on online content has increased significantly from 2001 to 2003 (Online Publishers Associations, 2002). A brief review of studies of WTP for digital network services shows that studies are conducted on WTP for broadband enabled services on the Internet available through computers (Sage Research, 2002; eBrain, 2002; Jupiter Research, 2003; Online Publishers Association, 2002; Makarem, 2001), digital services available through high definition television (HDTV) (The Leichtman Group, 2003), and mobile services (Coleago Consulting, 2002; Nokia, 2001; Ericsson, 2002; Forbes, 2003). In this article, the summary of the main findings for mobile services is based on a limited number of empirical studies. However, for 3G mobile services, entertainment services seem to have a high potential (Nokia, 2001), although there are some interest reported also for utility services as news delivery (Forbes 2003) and travel information (Driscoll and Associates, 2003). Also, for general broadband enabled services, several studies indicate that entertainment services are among the services with the largest market potential (Sage Research, 2002; Lyra Research, 2001; Online Publishers Association, 2002). Within this category, movies on demand (Sage Research, 2002; Myrio, 2003), concerts and cultural events (Sage Reserach, 2002), TV shows on demand (Sage Research, 2002; Myrio, 2003), and music download (Sage Research (2002) seem to be the services customers are willing to pay most for. For digital services, studies have focused age (Coleago Consulting, 2001; Myrio, 2003), probability of adoption (Colleago Consulting, 2001), stage of adoption (Nokia, 2001), uniqueness of service (Anstine, 2001; Edison Media Research, 2002), access to broadband (Jupiter Research, 2003), convenience (Molteni and Ordanini, 2003), and whether copies and file transferring is allowed (eBrain, 2002; Jupiter Research, 2003) as variables moderating customers' willingness to pay.

2 PURPOSE OF THE ARTICLE

Studies on WTP for services in digital networks have three typical characteristics. First, the methodology used to measure willingness to pay is typically based on surveys where the respondents hypothetically are asked how much they are willing to pay for a specified service (Coleago Consulting, 2001; Sage Research, 2002; Makarem, 2001; Leichtman Research Group, 2002). This means that the studies are not conducted in a real purchase situation, and the respondents are not given any specific incentives for revealing their true willingness to pay. This method has often been criticized for overestimating consumers' willingness to pay (Botelho and Pinto, 2002; Wertenbroch and Skiera, 2002). The first purpose of this article is therefore to discuss the importance of studying consumers' WTP for digital services in a more realistic context and with more realistic incentives.

Second, existing studies only marginally discuss willingness to pay as a theoretical construct (Ericsson, 2002; eBrain, 2002; Makarem, 2001; Edison Media Research, 2002; Nokia, 2001; Coleago Consulting, 2001; Leichtman Research Group, 2002). It is argued by Wertenbroch and Skiera (2002), that when using the survey method, application of a variety of willingness to pay measurement procedures should be used for cross validation purposes. Thus, supplementary theoretical perspectives from the nomological network of willingness to pay should be included to evaluate measures of consumers' willingness to pay for services in digital networks. Although some variables are revealed to moderate consumers willingness to pay for digital services, several other variables revealed to influence consumer behavior in general also seem to be relevant to include. Thus, the second purpose of the article is to highlight the importance of a broader theoretical manifold in WTP studies and to present a nomological network of WTP, including a discussion of potential moderating variables.

Third, the studies typically measure customers' willingness to pay for single-channel services (Sage Research, 2002; Makarem, 2001; Myrio, 2003). Often, a brand or a company is offering a service through more than one channel. However, we have not been able to find studies focusing willingness to pay for content in a multi-channel/cross-channel service compared to a single-channel service. Consequently, the third purpose of the article is to address the importance of focusing willingness to pay for cross-channel and multi-channel digital services.

2.1 Purpose 1: Context and incentives when studying willingness to pay

Four methods to estimate willingness to pay are described by Wertenbroch and Skiera (2002). The methods differ in their incentives to reveal true willingness to pay and simulation of actual point of purchase context. With the transaction method, respondents receive a fee before the estimation of WTP begins. They are then exposed to products where the price is manipulated, and they chose to buy the product with the manipulated price they perceive as "right". The method can be pure experimental or it can take place in a real purchase situation. The fee given to the respondents reduce the possibility that estimates are biased downward due to liquidity constraints. Thus, incentive compatibility should be considered relatively good when using the transaction method. By using the transaction method, only the pure experimental approach will be useful for revealing willingness to pay for products and services yet not on the market. However, in pure experimental studies, simulation of actual point of purchase context suffers, and the revealed willingness to pay may be biased.

One of the survey methods discussed by Wertenbroch and Skiera (2002) is based on contingent valuation of customers' WTP for a given product. This method could be used by an open ended approach, where customers are asked to state their WTP for a given product, or by a close-ended approach, where customers' choose whether to buy a product at given price alternatives or not. When using the contingent valuation method respondents are typically given a description of the product/service evaluated and the hypothetical circumstances under which it is made available. Thus, the WTP estimated by the respondents do not have any implications for the respondents, and the context is hypothetical. The survey method, therefore, scores relatively low both on incentives to reveal true WTP and on simulation of actual context.

The Vickrey auction (Wertenbroch and Skiera, 2002) is a sealed-bid auction. The incentive compatibility in this method is relatively good, because the respondent with the highest bid must buy the product. Thus, the WTP estimated by the respondents have implications for the respondents. However, there is a limited stock of products, so respondents compete with one another to

win the auction. The stock constraint used in Vicroys method contrasts the usually unrestricted supply of products in a real retail setting. The method is experimentally, and respondents meet and make decisions regarding willingness to pay in a research facility. Thus, simulation of actual point of purchase suffers when using the Vickrey auction method.

With the Becker, DeGroot, and Marschak's (BDM) (Wertenbroch and Skiera, 2002) method, respondents are told that the price of the given product is not yet set, and that we (the researchers) want to know the highest price the respondent is willing to pay for the product. When the respondent has reported his final WTP, he draws a price from an urn. This price will be the actual transaction price. Thus the actual transaction price is set randomly, and the actual transaction price is exogenous to respondents WTP. If the price is less or equal his WTP the respondents are offered the opportunity to buy the product for the price drawn from the urn (which is equal or lower than his WTP). If the price he draws is higher than his WTP, he will not be able to buy the product. The method is also based on "out-of-pocket transactions". Thus, the WTP estimated by the respondents has some implications for the respondents. By using BDM, customers' willingness to pay can be estimated at the point of purchase and thus vary according to purchasing context and competitive set. This possibility ensures point of purchase realism when using the BDM-method.

Table 1: Characteristics of the four methods.

Method	Incentive compatibility	Point-of-purchase realism
Transaction method	Good	Bad
Survey method	Bad	Bad
Vickrey auction method	Good	Bad
BDM method	Good	Good

The overview given in the table are somewhat simplified due to the fact that there are versions of the four methods described. For example, point of purchase realism for the transaction method depends on whether it is a pure experimental approach or a quasi-experimental approach.

The review of WTP studies for digital services shows that most of the studies are based on the survey method with a contingent valuation approach. Several studies have revealed that the use of the contingent valuation model overestimate customers' willingness to pay compared to customers' real willingness to pay (Botelho and Pinto, 2002; Wertenbroch and Siera, 2002).

2.2 Purpose 2: Nomological network of willingness to pay

Below we briefly review a few other theoretical perspectives related to WTP in order to establish the nomological network of willingness to pay.

2.2.1 Acceptance and adoption

In information systems research, two constructs are closely related to willingness to pay - adoption and acceptance. Three models stand out as the most widely applied to explain information systems and service adoption intentions. The technology acceptance model (TAM) (Davis, 1989) focuses on the cognitive and affective explanations of intention to use a specific technology or service. It includes five concepts - perceived user friendliness, perceived usefulness, attitudes towards use, intention to use and actual use. TAM may be seen as a special case of the TRA - theory of reasoned action (Fishbein and Ajzen, 1975). TRA includes four general concepts - behavioral attitudes, subjective norm, intention to use and actual use. In general TRA does not propose specific determinants of behavioral attitudes (attitudes towards use). However, in TRA these attitudes are composed of the belief that use leads to certain outcomes, and the evaluation of the desirability of these outcomes. The inclusion of subjective norm, however, represents an important additional social explanation when compared to TAM. The theory of planned behavior (TPB) was proposed as an extension of the theory of reasoned action to account for conditions where individuals do not have complete control over their behavior (Ajzen, 1991). This is particularly relevant to situations of service adoption under financial or other behavioral constraints. Perceived behavioral control - as included in TPB - reflects the internal and external constraints on behavior, and is directly related to both behavioral intention to use and actual use (Ajzen, 1991).

2.2.2 Marketing and consumer psychology research

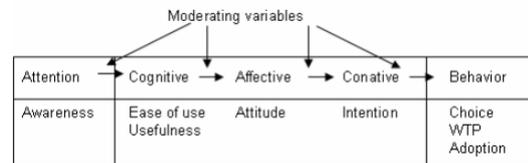
Constructs positively related to customers' willingness to pay in marketing and consumers' psychology is awareness, attitude, intention, and choice. Awareness reflects customers' awareness, their interest level, or their readiness for a service or a product (Schiffman and Kanuk, 1997). Peoples' consciousness, or awareness, about an object is found to influence customers' willingness to pay for the object positively (Vlosky, Ozanne and Fontenot, 1999). When it comes to attitude and intentions, theories as theory of reasoned action and theory of planned behaviour are also used in marketing. Attitudes and intentions are assumed to influence actual choice positively. Positive effects of attitude toward an object on willingness to pay for the object are revealed by e.g. Rio, Vazquez and Iglesias (2001).

2.2.3 Moderating variables

In the review of studies focusing WTP for digital services we reported several variables revealed to influence customers WTP for digital services. However, what is absent in this research is traditional moderating variables found to be of importance for customers behavior in many studies; such as consumers' channel

experience (Bruner and Kumar, 2000), product knowledge (Bei and Widdows, 1999), product expertise (Alba and Hutchinson, 1987), perceived risk (Dowling and Staelin, 1994), and product involvement (Bei and Widdows, 1999).

The nomological network is illustrated in figure 1.



By including several variables from the nomological network of WTP, measures of WTP can be validated against other measures related to WTP. Furthermore, as can be seen from figure 1, the variables included in the nomological network can be seen as a causal flow of psychological processes starting with product awareness, mediated by cognitive, affective, and conative mental processes, ending with a perception of WTP and a decision (choice) to buy/adopt the product/service or not. Thus, the variables included in the nomological network of WTP may also be used to explain the psychology behind the observed WTP.

2.3 Purpose 3: Willingness to pay for cross-channel and multi-channel services

Cross-channel services refer to services with identical content distributed in two or more channels. Multi-channel services are defined as services with supplementary or complementary content in additional channels. An example of a cross-channels service is when television companies give customers the possibility to watch television programs on their computer through the Internet at the same time as it is broadcasted on television. However, it is a multi-channel service when customers are given the opportunity to use their mobile phones to vote or make comments when watching a television show.

What is interesting in such a services context is whether access to content in a new channel has an influence on customers' willingness to pay for content in existing channels. One way to shed some light on the importance of this is to relate the discussion to whether cross-channel and multi-channel services are substitutes, supplements or complements to each other. If people start to watch television shows on their computers through the Internet in stead of watching on television, this is an example of channel substitution. If additional content related to a television show is offered through SMS or MMS without any changes in people watching the original show on television, this is an example of supplementary services. However, if these subscription services on SMS and MMS increase the interest in the television show, the additional channels are complementary to the original channel. In particular in

situations characterized by supplementary and complementary channels, willingness to pay becomes more complicated. When channels are supplementary (see e.g. Riel, Liljander and Jurriens, 2001), people are using a new channel in addition to the original channel to get access to the service. This does not influence their interest or use of the original channel, meaning that only the use of the new channel represents an increased interest or use of the service. Given that the new channel used is not free, this represents an increased willingness to pay for the service. In a complementary situation (Dutta-Bergman, 2004), the content offered in one channel influence the interest and use of content available in other channels. Thus, we have a situation characterized by channel synergies where perceived value of the content presented in one channel increase as a result of the additional content of the service offered in another channel. This increased value means people will be willing to pay more for the content distributed in the original channel in addition to their willingness to pay for the content distributed in the additional channel.

3 PROPOSITIONS FOR FUTURE RESEARCH

3.1 Context and incentives when studying willingness to pay

The measures of customers' willingness to pay for digital services are, due to the heavily use of surveys methods (contingent valuation approach), probably not very precise. Thus, an important issue in future research should be to develop more accurate methods to measure customers' willingness to pay for digital entertainment services. In particular, we have pointed to the importance of more realism in incentives and point of purchase context, as discussed by Wertenbroch and Skiera (2002). Thus, a main issue we suggest should be focused in future studies of WTP for services in digital networks is:

Future research 1: to deploy methods that includes more realism in incentives and point of purchase context.

3.2 Nomological network of willingness to pay

Most of the research on WTP for digital services is only marginally based on theoretical discussions and reflections. The ones who are have focused relatively narrowly on the WTP construct. One way to overcome this weakness is to build on the nomological network of willingness to pay. By doing this, measures from various theoretical perspectives can be used, and the results revealed from the various theoretical perspectives can be validated against each other – giving us more trustworthy results of customers' willingness to pay for-, accept-, adopt-, use-, etc. digital entertainment services. Therefore, one of the issues that should be focused in future research on customers' willingness to pay for digital entertainment services is:

Future research 2a: to deploy of a broader spectre of

methods and measures to reveal what we believe will be a more correct estimate of customers' willingness to pay for digital services.

In addition, the inclusion of variables in the nomological network of WTP makes it possible to explain the causality between variables in the nomological network of WTP - increasing our understanding of how to influence consumers' WTP. Future research should therefore;

Future research 2b: study the causality between variables in the nomological network of WTP.

The discussion presented in this article also reveals that relevant moderating variables found to influence consumers' behaviour in general are not included in studies of customers' WTP for digital services.

Future research 2c: Effects of potential moderating variables should therefore be included in future studies to get a more nuanced picture of willingness to pay for digital services in various customer segments.

3.3 Willingness to pay for cross-channel and multi-channel services

None of the reviewed studies of willingness to pay for digital services focus on willingness to pay for cross-channel services or multi-channel services. Several digital services are available in more than one channel. Thus, it will be of interest to study changes in willingness to pay for digital services in the main channel when the digital services are presented in additional channels as well. We therefore suggest that future research should;

Future research 3: study customers' willingness to pay for not only singles channel services but also cross-channel services and multi-channel services.

REFERENCES

- [1] Alba, J. W., and Hutchinson, J. W., "Dimensions of Consumer Expertise", *Journal of Consumer Research*, vol. 13, March, pp 411-454, 1987.
- [2] Ajzen, I., "The theory of planned behavior", *Organization Behavior and Human Decision Processes*, vol. 50, pp179-211, 1991.
- [3] Bei, L.-T. and Widdows, R., "Product Knowledge and Product Involvement as Moderators of the Effects of Information on Purchase Decisions: A Case Study Using the Perfect Information Frontier Approach", *The Journal of Consumer Affairs*, vol. 33, no. 1, pp165-186, 1999.
- [4] Botelho, A., and Pinto, L. C., "Hypothetical, Real, and Predicted Real Willingness to Pay in Open-Ended Surveys: Experimental results", *Applied Economics Letter*, vol. 9, pp993-996, 2002.
- [5] Bruner, G. C., and Kumar, A., "Web Commercials and Advertising Hierarchy-of-Effects",

- Journal of Advertising Research, vol. 40, nos. 1&2, pp35-42, 2000.
- [6] Coleago Consulting, "Demonstrating Demand for Mobile Internet Access", Author; Stefan Zehle, www.coleago.com, 2001.
- [7] Davis, F.D., "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology", MIS Quarterly, vol. 13, pp319-340, 1989.
- [8] Dowling, G. R., and Staelin, R., "A Model of Perceived Risk and Intended Risk-handling Activity", Journal of Consumer Research, vol. 21, June, pp119-134, 1994.
- [9] Driscoll & Associates, "Commercial Telematics Systems and Service Study", www.cjdriscoll.com, 2003.
- [10] Dutta-Bergman, M. J., "Complementary in Consumption of News Types Across Traditional and New Media", Journal of Broadcasting and Electronic Media, vol. 48, no. 1, pp41-60, 2004.
- [11] eBrain, "Digital Download and Home Recording Issues", October 2002.
- [12] Edison Media Research, "Advertising vs. Subscription – Which Streaming Model Will Win?", Arbitron Inc, Edison Media Research, 2002.
- [13] Ericsson, "Mobitex Business Guide", Ericsson AB, www.ericsson.com/mobitex, 2002.
- [14] Fishbein M. and Ajzen, I., Belief, attitude, intention and behavior: an introduction to theory and research, Addison-Wesley, Reading MA, 1975.
- [15] Forbes, "Forbes Study Finds Willingness to Pay for News Via Cellphone", B to B, vol. 88, no. 1, p13, 2003.
- [16] Jupiter Research, "Downloaders Will Pay for Ownership", www.cyberatlas.internet.com/markets/retailing/article, 2003.
- [17] Leichtman Research Group, "Consumers Willingness to Pay for Subscription Video-on-Demand", www.leichtmanresearch.com, 2002.
- [18] Leichtman Research Group HDTV, "Consumer Awareness, Interest and Intent to Purchase", www.leichtmanresearch.com, 2003.
- [19] Lyra Research, "Online Content: Gotta Pay to Play?", Biz Report, www.bizreport.com/article.php?id=1495, 2001.
- [20] Makarem, R., "Ways and Means. Identifying winning on-line applications over multiple platforms", Cap Gemini Ernst & Young, Strategic Research Lab, 2001.
- [21] Molteni, L., and Ordanini, A., "Consumption patterns, digital technology and music downloading", Long Range Planning, vol. 36, pp389-406, 2003.
- [22] Myrio, "The Value of Middleware. Taking IPTV from Headend to Home", Technical Paper, www.myrio.com, 2003.
- [23] Nokia, "3G Market Research. Mobile Entertainment: An End User Perspective. Key Learnings from a Global Market Research", www.nokia.com/3G, 2001.
- [24] Online Publishers Association, "Online Paid Content. U.S. Market Spending Report", http://www.online-publishers.org/opa_paid_content_report_122002_final.pdf, 2002.
- [25] Riel, A. C. R., Liljander, V., and Jurriëns, P., "Exploring Consumer Evaluations of E-Services: A Portal Site", International Journal of Service Industry Management, vol. 14, no. 4, pp359-377, 2001.
- [26] Rio, A. D. del, Vazquez, R., and Iglesias, V., "The Effects of Brand Associations on Consumer Response", Journal of Consumer Marketing, vol. 18, no. 5, pp410-425, 2001.
- [27] Sage Research, "Customers at the Gate: Mounting Demand for Broadband-Enabled Services", www.sageresearch.com, 2002.
- [28] Schiffman, L. G., and Kanuk, L. L. Consumer Behavior, Prentice-Hall, Upper Saddle River, New Jersey, 1997.
- [29] Vlosky, R. P., Ozanne, L. K., and Fontenot, R. J., "A Conceptual Model of US Consumer Willingness-To-Pay for Environmentally Certified Wood Products", The Journal of Consumer Marketing, vol. 16, no. 2, pp122-, 1999.
- [30] Wertenbroch, K., and Skiera, B., "Measuring Consumers' Willingness to Pay at the Point of Purchase", Journal of Marketing Research, vol. 39, May, pp 228-241, 2002.

Working Mechanism and Structure of Customer Services Support System

Qingguo Wang, Shuqin Cai, Yuhua Zhou, Kaichun Liang

School of Management, Huazhong Univ. of Sci. & Tech., Wuhan 430074, China

wqg1997@163.com

ABSTRACT

On the basis of analyzing customer services of manufacturing enterprises and these supporting requires, using the method of discrete system simulating, the paper provides the classifications of customer services suitable information technology supporting, brings forward the working mechanisms suitable to the supporting system of different kinds of customer services and the relevant system structure.

Keywords: customer service (CS), customer service support system (CSSS), working mechanism, system structure

1. INTRODUCTION

To manufacturing company, with the fierceness of the market competition, customer service of enterprise faces all kinds of presses such as the increase of working load, serve standard raising, resource limited and fund shortage,^[1] and customer service bears dynamics, complexity, timeliness and outburst of time and space, so enterprise requires customer service supported by information technology which can be realized with low cost, effective resource allocation and fast and automatic reactivity. The domestic and international relevant scholars have studied in regarding customer service and its supporting technologies with information technology^[2-9], the typical applications include mainly: customer care, dispute disposing, order following, field service, and database of problem and solution, etc. Many kinds of new methods of customer service such as Web, VoIP, Email, Live Chat, Virtual Reps have appeared. But the applications of the technology are mainly the support to the concrete business of customer service and have supported the development of enterprise's customer service to a certain extent, but set foot in little in decision of enterprises service resource scheduling. The problems such as rational arrangement and managing enterprises service resources, integrating the numerous service methods and formulation of the service tactics are that enterprises are badly in need of solution at present. Though customer service has many kinds of different classification already,^[10,11] it can't totally respond to the request of customer service supporting with information technology yet. So it is very necessary to study the classification of customer service in more pertinence, which can help to provide different and more effective disposing mechanisms for all kinds of customer services.

Customer service supporting system is a part of customer relationship management system backstage supporter their system, supports and restricts whole customer relationship construction and service support operation, and plays a key role in the system, so its working mechanism and system structure become hotspot in theory and practice to study. Systematic simulation has already been applied to numerous

systems such as economic system, business system, and logistics management system extensively. By the look of resource competition, customer service is a kind of question of lining up, and dispersed incident systematic simulation is a kind of auxiliary administrative skill and can preferably solve this problem^[12,13]. Because of the characteristics of customer services, the study of customer service not only focuses on simulating experiences and suitable service strategies, also includes the customer service support system that integrates the testing study and practicing support.

Taking manufacturing enterprise as the background, this paper provides a kind of new customer service classification based on discussing influence factors of customer service, designs system model of event-driven customer service using systematic simulation technology, provides system structure and the corresponding systematic module to provide guidance for development of enterprise's customer service, distribute the service resources of enterprises rationally in the course of implementing and improve the customer satisfaction of enterprises.

2. CUSTOMER SERVICE CLASSIFICATION SUITABLE TO CSSS

The customer service is a very extensive concept, does not confine to narrowly-defined ranges such as traditional selling service and after-sale service and should include these activities, such as advertisement, free trial, presentation and investigation, solicitude and support to customers which get in touch with enterprise, involving the whole cycle of selling products. So the customer service is not only the service request responding the customer passively briefly, should also offer customer service voluntarily.

2.1 Influencing Factors of Customer Service

The service time variable not only points the implementation time of the customer service but also includes the makeup-time and latency time of service. As latency time correlates with the customer service tactics of enterprises and customer service resource and

does not have direct relation with customer service, only implementation time and make-up-time of the customer service are considered here. In the customer services that enterprises offer, some are finished once, otherwise, some are implemented many times stage by stage, where the service time of this kind of customer service is made up of implementation time of a lot of customer services. So service time T can be described as following:

$$T = f_1(n), \quad (1)$$

where n denotes stage degree of service. When n is equal to one, T is short, and T is long when n is bigger than one.

Service costs include only labour cost l , material cost m and opportunity cost. Labour cost depends on the service provider and shows significant degree of customer attendants in enterprises. The material cost is the value of the corresponding material while offering this kind of customer service by enterprises, if maintaining the three guarantees products, enterprises should offer the corresponding spare part to replace the trouble one. When a certain customer service needs to consume material cost, not only raise the service costs of enterprises, may take more makeup-time for planning the material. The opportunity cost means possible losses caused by that enterprises appoints competent technology and management attendants to implement one requiring the low customer service to engineering level. Because the attendants with competent technology are limited in quantity, so this kind of chance loss may cause another customer service expected much to engineering level not to get prompt treatment. So service costs can be described as the following:

$$C = f_2(l, m), \quad (2)$$

where l is low without m , C is low, and C is high when l is high or there is m .

Different customers have different customers' value to enterprises. If enterprises do not distinguish the customer according to these value, it will cause that some customers' actual contribution to enterprise are still not enough to remedy the input of enterprises for them, otherwise, some customers are influenced as investment to them is insufficient. According to customer value, all the customers can be classified into important customers and ordinary customers shown by G . There are present customers in the important customers, and there are potential customers too. The important customer enjoys enterprise's customized customer service, and enterprises offer ordinary customer service for ordinary customer.

2.2 Classification of Customer services

Considering above-mentioned influence factors synthetically, customer service classification Y can be

defined as follows:

$$Y = f_3(T, C, G) \quad (3)$$

The customer services of manufacturing enterprises can be divided into event type, activity type and process type three kinds, according to T, C, G , and the requires, feasibility and simplicity realized of information supporting technologies to customer services, showing as table 1.

Table 1 Value of Customer Service Classification

Y	T	C	G
Event Type	Short	Low	Ordinary
Activity Type	-	High	Ordinary
Activity Type	Long	Low	Ordinary
Process Type	-	-	Important

Event type customer service is a kind of customer service that possesses of short service time, low service cost and ordinary customer, which is one stage service. The realization of event type customer service can be regarded as the array of event serials made up of beginning event and ending event.

Activity type customer service is a kind of customer service that possesses of long service time or high service cost and ordinary customer, which is multi-stages service, and the beginning of subsequent service decides by the foregoing service. Activity type customer service accords with the definition of activity in activity scan simulation method, so it can be disposed as an activity.

Process type customer service has nothing with two factors of service time and service cost, only decided by customer types, whose customer is important customer. It is not a certain concrete customer service, but a kind of customer service tactics of enterprises, it covers the customer's whole life cycle and is a service package including lots of services for the important customers. Each service package is different from others as each of them faces a certain important customer. So process type customer service is made up of event type customer services and activity type customer services disposed as event and activity respectively. A process type customer service including these events and activities is disposed as a process.

Three kinds of customer services have different service costs, among them the incident type has the lowest customer service, the course type is the highest, the differentiation of this kind of service cost contributes to controlling enterprise's service cost, and the enterprise can provide different customer services to different customers discriminatively. The serving objects of three kinds of customer services are different, those of the event type customer service and activity type customer service are ordinary customers, those of process type

customer service is important customers, which contributes to guaranteeing enterprise's main profit source and satisfies the service of the important customer. From the point of information technical support, above-mentioned categorized methods can help enterprises to implement customer's service support system based on information technology better.

3. WORKING MECHANISMS OF CSSS

Systematic simulation is a course of solving the dynamic system model with number value method, which understands and predicts systematic behavior through the experiment of a realistic system or the system model that plans to be set up. Event step method in dispersed incident systematic simulation is often used to simulate enterprise management system, which is divided into three kinds of different simulation working mechanism: event-driven method, activity-scanned method and process-switched method.

3.1 Working Mechanisms of CSSS used by Event-driven Method

Event-driven method is described in Ref. 12 as: Operations of system are made up by a series of emergence of events according to time, courses that take place inside are regarded one array of events and future event list (FEL) is adopted to confirm the order of events according to each time that events happen. The incident of time take place first is chosen from FEL, each time the most necessitous event is disposed, simulation clock is advanced to here incident time at the same time, then the corresponding event subprogram is carried out, the corresponding systematic state of changes, and any condition is tested and happened in the incident subprogram. After an event is dealt with, it retreats from FEL. The operation of the event-driven simulation method includes two stages: (1) Time scanning, to carry out two operations: scanning future incident form FEL to confirm next time that event happens, and moving e forward clock to present time. (2) Event execution, to transfer the corresponding event subprogram. Carry out two these repeatedly, until the fact that fixed simulation end conditions are satisfied.

In three kinds of customer services, the event type customer service is made up of service beginning event and service ending event, and the emergence course of it also composes of the array of these two events above. The activity type customer service can be regarded as a lot of associations of event type customer services, it also composes of service beginning event and service ending event too, but a different one is that the end of anterior events can touch off the beginning of the next stage. The process type customer service is the service package made up of event type customer services and activity type customer services. So Event-driven method can be adopted to support tactics and simulation implement of the three kinds of customer service above.

3.2 Working Mechanisms of CSSS used by Activity-Scanned Method

Activity-scanned method is the simulating method based on activity scanning. Beginning events and ending events of are initialized by terms that the activity stipulate, and not prearranged by the model. In three customer services, event type customer service does not adapt to activity-scanned method because it is one stage service. Activity type customer service is made up of lots of stages where subsequent service is decided by anterior service, so it is suitable to activity-scanned method. Process type customer service includes activity type customer services, so it can be disposed using activity-scanned method.

3.3 Working Mechanisms of CSSS used by Process-switched Method

Process-switched method is the simulating method based on processes that compose of the array of events and some activities. Event type customer services and activity type customer services are unsuitable to process-switched method because of their characteristics. Process type customer service is made up of event type customer services and activity type customer services, so it can be disposed using activity-scanned method.

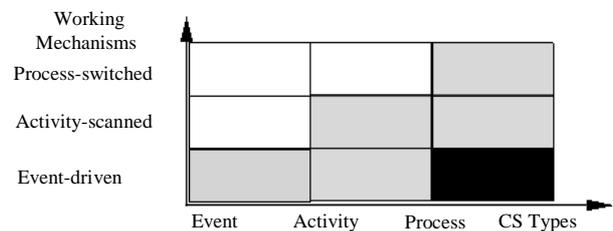


Fig. 1 Working Mechanisms suitable to each kind of CS

Fig.1 shows the different working mechanisms supporting the three kinds of customer services in customer service support system.

4. STRUCTURE AND REALIZATIONS OF CSSS

4.1 Main Data Structures of CSSS

The static information of system consists of customer information, customer service information, the detail of activity type customer service, the detail of process type customer service, distant time weight, distance weight, balanceable guideline weight, service attendants' information and city geographical information. Here, customer information is the relative information to customer weight, and not customer dossier in customer relationship management. Customer service information records all the services developed by the enterprise. The detail of activity type customer service records the information of each stage services in activity type customer service. The detail of process type customer

service records all the service items of important customers served in process type customer service, which includes the following word sections: customer number, product number, service number and prospective service time, where prospective service time denotes prospective time of each service item. Distant time weight records every weight that increases after certain time interval to count. Distance weight records every weight that increases after certain distance interval to count. Balanceable guideline weight records the weight in total PRI of each index. Service attendants' information records the personal information and technology level information of all the attendants. City geographical information records the longitude and latitude of all over the national cities, and the grain size reaches the county.

The dynamic information of system consists of the information in the following tables: present event list, future event list and customer latency time list. Present event list records present customer services serving, future event list records waiting customer record having been in service queue, and customer latency time list records the latency time taking in each serving.

Service tactics storehouse mainly records service tactics used by the enterprise and their purposes. In system, service tactics represents as variation of all the weights, and their purposes are mainly evaluated by the average of latency time. In service tactics storehouse, there are three kinds of average of latency time: the average of latency time of important customers, the average of latency time of ordinary customers and the average of latency time of all the customers, which are used to judge the influence of service tactics on the important customers, ordinary customers and the whole customer cluster of the enterprise. These information in service tactics storehouse not only are an evaluation to former service tactics, also use for reference to drawing new service tactics.

There are three kinds of events: happening event of new service need, service ending event and latest approaching clock event, to vary the systematic state. Happening event of new service need denotes the new service need being from customers or new customer service practiced by the enterprise, whose direct influence is to add new record in FEL. Service ending event denotes that the serving customer service has completed, whose direct influence is this record of customer service will be deleted from FEL. Because not all the waiting service records need to be served immediately, the latest approaching clock event is set to dispose them. When latest approaching clock event is touched off, system transfers directly the calculating rules of entering FEL, inserts suitable record into FEL and make it be served.

4.2 Function and Structure of Event-driven CSSS

Customer service support system is one of the back platform subsystems of customer relationship management system. It supports customer service and the practice of customer service, makes it be the profits center by distributing customer service resources and decreasing costs. Because of the concrete difference in business treatment of customer service, as an event-driven customer service system in common use, it does not participate in the actual treatment of enterprise's customer's service items, but regarded as a middle bridge between customer identifying system and enterprise's customer service process system. The event-driven customer service system receives service needs of customers issued by the intellectual agent's system of the customer service, optimizes these needs, distributes the service resources of enterprise in reason, and submits them to enterprise's customer's service item process system. So its main function is the support to decision-making and practices. Fig. 2 shows the information flows of customer service support system, and Fig. 3 describes the structure of the system.

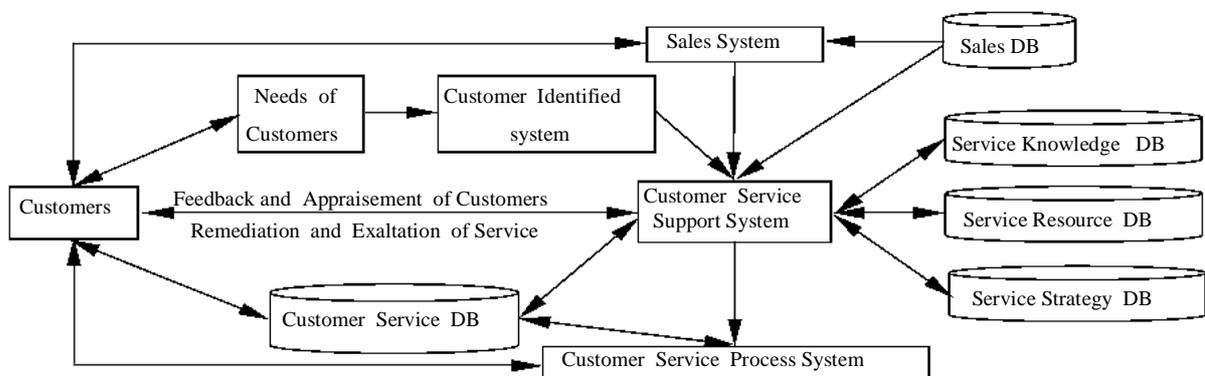


Fig. 2 Information Flows of Customer Service Support System

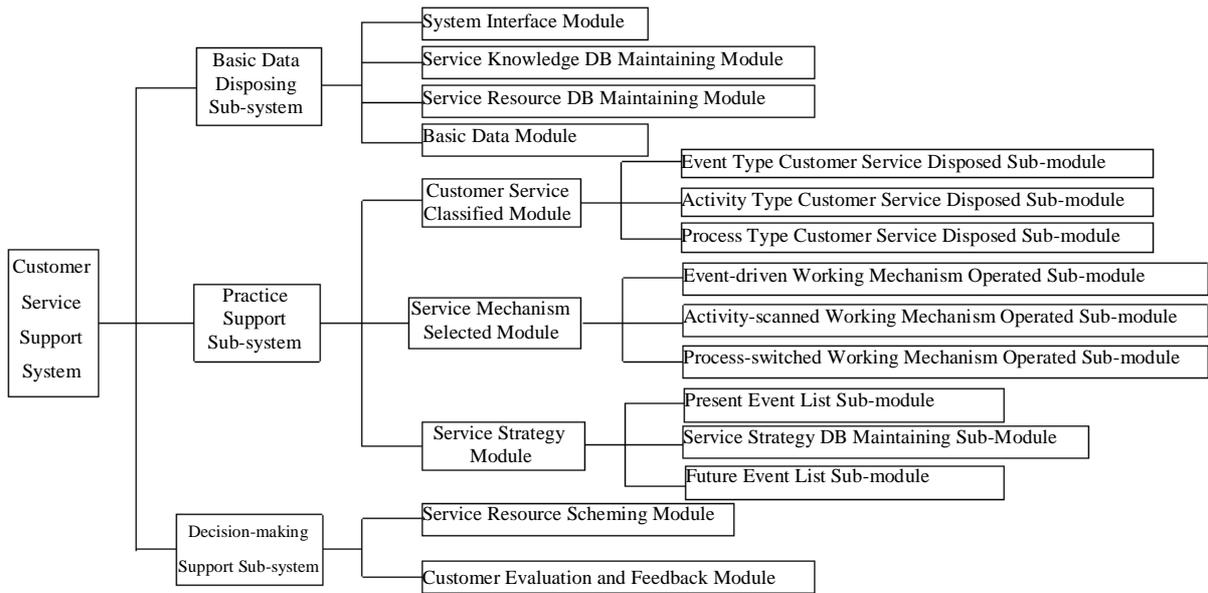


Fig. 3 Structure of Customer Service Support system

4.3 Realization of Event-driven Working Mechanism

Event-customer service is not a concrete customer service, it is a kind of customer service with the same working mechanism, though the service content of each kind of customer service has very great differences, they possess the characteristics of event-customer service such as the short service time, low service costs and general customers, so their way of taking up enterprises service resources is the same. The event type customer services accords with the so-called a M / G / 1 system^[14] ^[15] ^[16], namely a queuing model where the flows of customer reaching are the simplest flows, service time is at random and it possess one information desk. So the realization course of customer service A can be simulated using M / G / 1 system.

The simulative models of event-driven method include defining entity and entity set, defining activity and event, and describing event disposal. Here, the entities compose of a lot of customers and an information desk, the entity set is customer queue, the activities include waiting, being served and falling out the queue, the events compose of reaching information desk named event 1 and going away from information desk named event 2. The disposing procedures of event 1 and event 2 can refer the relative reference. Fig.4 describes the disposing procedure of event-driven working mechanism for event-customer service.

The simulating model of activity-customer service is similar with that of event type customer service except that the customer is served by service B₁ and service B₂ in turn in activity type customer service. We can suppose that customers arrive at $t_1, t_2, \dots, t_n, \dots, t_{n+m}, \dots$, where we can think that a customer is served by service B₁ at t_n and by service B₂ at t_{n+m} , the service times are independent and the twice arrives of a customer are

thought as the arrives of two customers. The flows of customers reaching are the simplest flows, so activity type customer services can also be simulated using M / G / 1 system. There are three kinds of events: reaching information desk named event 1, producing condition after service B₁ named event 2 and going away from information desk named event 3, where the disposing procedure of event 1 and event 3 is similar with those of event-customer service, and other disposing procedures are omitted limit to space.

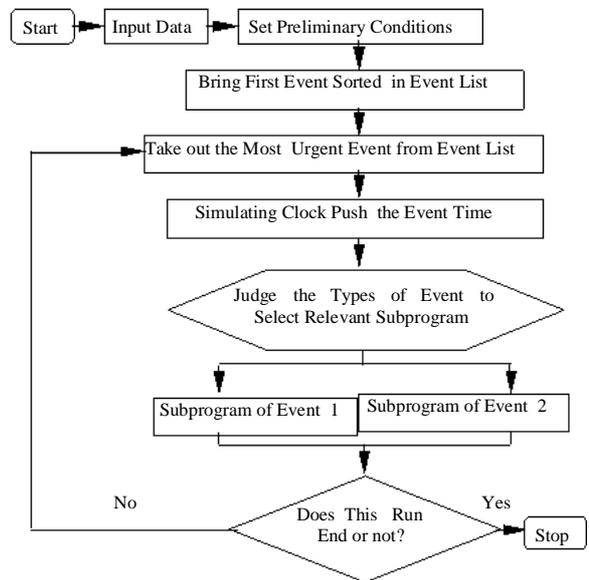


Fig. 4 Procedure of Event-driven Working Mechanism for Event-customer Service

Because the process-customer service is a service package made up of event type customer services A and activity-customer services B, we can adopt the procedures of event type customer service and activity-customer service in simulation of process-customer service, whose simulating model and

events are made up of the models and events of event type customer services and activity type customer services. The relative models and disposing procedures are also omitted.

5. CONCLUSION

According to the three variables: service time, service cost and customer type, this paper provides the three types of customer services: event type customer service, activity type customer service and process type customer service. Simulating the procedures of these customer services using event-driven method and establishing the system structure of customer service support system, which help enterprises to optimize the procedure of customer service, raise the efficiency, lower costs, reduce service accident, and this customer service support system model has oneself ability of appraising and adjustment.

Of course, this paper has laid particular emphasis on studying event-driven working mechanism, and there is the research to activity-scanned working mechanism and process-switched working mechanism remaining to continue carrying on deeply. At the same time because different working mechanism have different influence on way and result in which the customer service is implemented, it is necessary to compare the result of adopting different working mechanism of three kinds of customer services, find out the most suitable working mechanism of each kind of customer service, then set up system in common use. As regards technological aspect, information technology adopted in different working mechanism need to study further. At aspect that system implement, customer service support system and other data of system share also await to study and discuss.

ACKNOWLEDGEMENT

This work was supported by national nature foundation of China 70271030 and research foundation of graduate school of Huazhong Univ. of Sci. & Tech 2002.

REFERENCES

- [1]Greg Gianforte, "The Future of Customer Service: The Road to Top-Line Impact", <http://www.rightnow.com>, 2003
- [2]Berkley B J, Gupta A, "Improving Service Quality with Information Technology", *International Journal of Information Management*, Vol. 14, No. 2, pp 109-121, 1994.
- [3]Domegan C T, "The Adoption of Information Technology in Customer Service", *Journal of Marketing*, Vol. 60, No. 6, pp 52-69, 1996.
- [4]Wang Chunxiao, Wen Biyan, "Internet- A New Era of Customer Service", *Science Technology and Management*, No. 1, pp12-15, 2001.
- [5]Zhang Ning, Ye Xiaojun et al, "Design and Implementation of the Design-support System in a Customer Service Management System", *Computer Engineering and Application*, Vol. 37, No. 16, pp107-109, 2001.
- [6]Lu Jianfeng, Zhang Hao et al, "The Construction of A Tele-customers Engineering Support System and Its Integration with PDM System CESS", *Manufacturing Technology and Machine*, No. 2, pp12-14, 2002.
- [7]Jiang Xianxin, Zou Hua, Wang Bo, "A Net Meeting-Based Integration of Call Center and Web Scheme", *Journal of Beijing University Posts and Telecommunications*, Vol. 23, No. 1, pp92-95, 2001.
- [8]Guo Xiurong, Liu Jidong, "Design and Application of Data & Voice Information Support System for Shandong Electric Power Customer Service Center", *Electric Power*, Vol. 29, No. 2, pp 63-66, 2002.
- [9]Luo Cheng, Chen Xuerong, Zhang Shaojun, "Application Study on VoIP Technology in 185 Customer Service Center", *Study on Posts*, Vol. 18, No. 2, pp27-28, 2002.
- [10]Wang Qingguo, Cai Shuqin, "Study on the Classification of Customer Service", *Management Review*, Vol. 16, No. 3, pp19-24, 2004.
- [11]Cai Shuqin, Wang Qingguo, Tang Yunfei, "Summary of Customer Relationship Management and Customer Services", *Forecast*, Vol. 23, No. 5, pp55-60, 2004.
- [12]Huang Wei-wei, "Method and Application of Management System Simulation", Beijing, Renmin University of China Press, 1991.
- [13]Shi Qing, "Stochastic System of Management", Beijing, China Machine Press, 1994.
- [14]B Krishna Kumar, A Vijayakumar, D Arivudainambi, "An M/G/1 Retrial Queuing System with Two-phase Service and Preemptive Resume", *Annals of Operations Research*, Vol. 113, No. 1, pp61-79, 2002.
- [15]Kailash C. Madan, "An M/G/1 queue with second optional service", *Queueing Systems*, Vol. 34, No. 4, pp37-46, 2000.
- [16]Onno J. Boxma, Tetsuya Takine, "The M/G/1 FIFO Queue with Several Customer Classes", *Queueing Systems*, Vol. 45, No. 3, pp185-190, 2003.

A Component-based Framework for Distributed Business Simulations in E-Business Environments

Alexander HAMPEL¹, Edward W.N. BERNROIDER²

¹ Department of Computer Science and Business Informatics, University of Vienna, Austria
alexander.hampel@univie.ac.at

² Department of Information Business, Vienna University of Economics and B.A., Austria
edward.bernroider@wu-wien.ac.at

ABSTRACT

Simulations preserve the knowledge of complex dynamic systems and consequently transfer the knowledge of the cohesions of its elements to a specified target group. As the progress in information technology and therefore the dynamic e-business driven economy adapts even faster to the business demands, new ways to preserve this growing amount of knowledge have to be found. This paper presents an extensible business simulation framework which is realized as a component-based distributed Java Version 2 Enterprise Edition (J2EE) architecture. The framework aspires to offer an extensible and domain independent simulation environment which ensures the return of investment in the sense of implementing this framework once and extending it to the future requirements of diverse domains in e-business. The system architecture follows the requirements in offering distributed deployment of its components on highly standardized level by nevertheless staying vendor independent. The architecture itself was developed by model driven architecture (MDA)-conform software engineering methods using best of breed design patterns composed to a flexible micro-architecture which possess import facilities for simulation entities (business objects) and (business) processes from e-business solutions. Combining the features of the framework, the layered pattern driven micro-architecture, and the distributed J2EE architecture, the postulated knowledge transfer from rapid changes in e-business can be realized.

Keywords: business simulation, framework, architecture, J2EE, e-learning, e-business

1. INTRODUCTION

Business simulations play a major topic in industry and higher education for more than 40 years. They preserve the knowledge of complex dynamic systems by modeling the cohesions and dependencies between the simulations' objects and transfer the modeled information by simulating its dynamic aspect on computer systems [1, 18]. In that way the knowledge about the behavior of the dynamic system is transferred to the specified target group [21]. Depending on what knowledge should be preserved, the cohesions of the system for the audience the simulation may differ. These advantages were identified by the American Management Association by developing its Top Management Decision Simulation in 1957. From 1957 till 1962, 89 different business simulations were developed by the industry, universities and other organizations [24]. In 1965 an accounting and book keeping simulation was developed to show the dynamic dependencies between the related records. Four types of simulation models were identified: strategic planning, internal planning, training program and research [19]. As a lack of computer systems these days, simulations were limited to non-executable implementations without having the ability to vary its modeled cohesions. As computer systems grew to solve more and more automated processes the awareness of linking these systems rose. The needs of integrating data and interconnecting networks (applications and systems) in business and industry was examined in depth to develop an open communication architecture [5, 29] and further

more frameworks of for holistic information systems architectures were defined [33]. Middleware was and is the key player in integrating distributes systems starting 1994 with OMG's object-oriented open CORBA standard: a vendor, platform and program language independent specification [4, 15]. Integration approaches have widen the standardization process and brought new concepts of component ware, integration via web services [3, 7] and enterprise application integration (EAI) solutions for integration on the levels of data, application and processes [31, 23]. All the mentioned integration efforts have led to the new dynamic area of the e-business driven economy which is able to adapt even faster to the business demands [20]. As a matter of fast changes in e-Business life-cycle of virtual enterprises, new ways to preserve this growing amount of business and process knowledge, and transfer of the dynamic content have to be found. The following simulation framework is based on our very recent research project.

2. BUSINESS SIMULATION FRAMEWORK

To set-up a business simulation without a framework on the meta-level would end up in a development without extensibility to adapt neither to new demands of the implemented business simulation nor the application of new simulation domains. These major disadvantages of having no continuous life-cycle for simulation systems led us to the development of a simulation framework to ensure a maximum of flexibility for the architecture applied as a corporate simulation and e-learning

instrument. The following top-level requirements were defined for the framework:

- extensible to adapt to new demands
- enable distributed deployment
- simulation domain independence
- ability for import simulation objects, (business) processes and (business) rules from existing e-business or enterprise resource planning (ERP) systems

In asynchronous e-learning systems the reuse of implementation as extensible frameworks [6] is more common as in the area of synchronous e-learning solutions like business simulations. Models and implementations of business simulations are custom tailored to certain needs without developing a general simulation framework as [25] shows for the segment of production management. For our framework to comply with the stated requirements above is divided into the following dimensions of which the components of the simulation framework will be derived:

2.1 Simulation Client and Server Environment for Distributed Deployment

Monolithic architectures, as stand-alone database applications [18], bring along disadvantages which led to client-server architectures in distributed systems. As the simulation framework is deployed in distributed environment client and server components have to communicate via TCP/IP. All used components need to get access and be accessed by standard protocols using a multi-layered architecture. Further there must be the ability for a number of simulation systems to be interconnected.

2.2 Access with HTML-Browser and XML-based Client Applications

As distributed multi-layered simulation framework there has to be access to the server components in all expedient ways. HTML-browser access the simulation components with low featured graphical requirements like administration of the simulation system itself or registration for participants. Is a more featured graphical representation for providing a complex user interface for the simulation clients required, access with client applications and the use of XML communication has to be provided.

2.3 Domain and Domain Independent Components

To enable the framework to be simulation independent applicable there has to be a distinction between domain independent simulation facilities like communication services or operation figures and specific domain dependent components like the used simulation objects.

2.4 Production and Administration Environment for Separate Deployment and Assuring Knowledge Preservation in a Repository

Offering a domain independent open simulation framework, there has to be a central repository to save all modeled knowledge. As models of business simulations can be continuously extended while in use and simulations of different domains can be applied to the framework, the global repository plays an important role. This is an important feature for a corporate group of companies which wish to install the implementation of our simulation framework once and use an instance of the simulation framework for each subsidiary for different domains – preserving a holistic all-in-one mapping. International e-learning communities as university undergraduate schools could develop simulation models for different subjects and share the simulation models using the framework.

2.5 Synchronous and Asynchronous Communication Facility; Simulation Inherent and External Communication Facilities

Communication between the simulation parties and between interconnected simulation systems plays a crucial role in a distributed environment. Therefore synchronous communication facilities are needed for the simulation parties to chat or to send instructions between clients and the server infrastructure. Asynchronous communication is needed for all long-run activities where no immediate feedback is necessary which applies mostly for communication with external resources, e.g. feedback for the user registration process which is realized via SMTP.

2.6 Vendor Independence and Import Facilities from E-Business or EAI Solutions

To ensure a long life-cycle of the simulation framework vendor independence is an important issue. Some vendors use frequently update strategies of their products to keep the turn over high, to change their underlying architecture and APIs to stay in a proprietary way [9]. To adapt to new demands of the fast changing e-business, the simulation framework has to offer import facilities for simulation objects (business objects), (business) processes and rules, and from existing systems. As e-business solutions are set-up of different applications running on legacy systems, ERP systems or enterprise application integration (EAI) solutions, the import of these systems using standardized interfaces reduces the effort of creating new business simulations.

2.7 MDA-conform software engineering methods

As scientific long-term projects [18] and a large number of projects in the information technology have shown, the lack of using iterative and incremental development

using different standardized protocols. HTML-browsers use HTTP/HTTPS, Java client applications use Java-RMI and SOAP/XML is implemented to extend the client type to Macromedia Flash applications [16] and other remote clients.

3.3 Domain Dependent and Domain Independent Components

Domain *dependent* components are business simulation specific objects which are realized as Business Objects [2, pp374] and designed as Composite Entity [2, pp391] as described in the J2EE [12] architecture below. Business objects are highly correlated with the import facilities of external systems and the simulation processes in this framework. As business objects unite business data and logic, they are on the one hand the key for mapping real-world or e-business objects to simulation objects in the way that they are the smallest entity, and no matter how the data is stored in the external system (e.g. split up into different tables), a composite view of the aggregated data is used as Composite Entity. The business logic stored as business processes on external systems and can be imported too. The business objects are aggregated to subsystem components and encapsulated on a higher abstraction layer. The import with direct access to the external systems (e-business, EAI solutions or ERP systems) using J2EE specifications is described later on. Export facilities are not provided in this framework nor the use of Learning Object Metadata [11] for the internal persistence layer as these features are no central issue in our research project.

Domain *independent* components are independent of the business simulation like the system boot-up process, loading, creating, importing and editing of simulations or the administration of the simulation environment. Further the communication objects and the accounting including operating figures to evaluate simulations are domain *independent*. They are on the other hand the key to the dynamics of business simulations: processes belonging to the business object are encapsulated as object's method and processes spanning various business objects are aggregated to use cases and combined as Application Service which represents the domain *independent* simulation engine.

The access to the persistence layer is managed by the J2EE application server [12] and non-DB sources by Data Access Objects [2, pp462]. Data Access Objects abstract and encapsulate the access to the persistence layer in the way that business objects do not have to have information about the persistence store and can transparently access all kinds of data like XML feeds or flat files.

3.4 Production and Administration Environment for Separate Deployment and Assuring Knowledge Preservation in a Repository

The framework is designed to be used as

multi-deployment implementation with a distinction between the production systems and the administration system containing the global repository (Figure 1). As there is the need of a central repository to back-up the simulation objects, the business processes and to preserve the modeled dependencies of all production systems, it is further used for documentation purposes. The set-up of new simulation models and the transfer to other simulation instances of this simulation framework is managed by the administration environment.

3.5 Synchronous and Asynchronous Communication Facility; Simulation Inherent and External Communication Facilities

To enable simulation inherent client to client, client to server communication, and external communication facilities, the communication component is part of the domain *independent* simulation components. As there is a multi-view to the components of the business simulation framework and the micro-architecture, these views overlap. Synchronous and asynchronous system communication is realized with the standard platform API. To summarize at this point for the J2EE architecture: Java-RMI is used for object calls, SOAP/XML for client-server and server-server communication and SMTP for e-mail.

3.6 Vendor Independence and Import Facilities from E-Business or EAI Solutions

Vendor independent development is realized by using a J2EE architecture as it is vendor, operating system and hardware independent. Depending on the application demands the simulation framework can be deployed on small, medium or large scale systems, standalone or clustered. The distributed J2EE architecture offers a connector architecture (JCA) [13] to enable all major platforms and vendors to provide their own resource adapters or implement the needed JCA connector with minimal interface programming themselves. Further as stated in the requirements for the framework, the only way to deliver an extensible system to adapt to new demands is to lengthen the life-cycle of the application by using sustainable and open standards. Our simulation framework therefore offers import facilities using JCA or indirect import of any desired semantics using XML or CSV format.

4. DISTRIBUTED J2EE ARCHITECTURE AND ITS IMPLEMENTATION

The sections above described the requirements, the deployment and the micro-architecture (both PIMs) of the framework. Following the MDA the next step towards the implementation is the transfer of the PIM to a Platform Specific Model (PSM) [22] to ensure the system architecture keeps the requirements: extensible, distributed deployment, simulation domain independent

and import capabilities. The chosen J2EE [12] platform and its implementation specific topics follow.

4.1 Basic Rules for Transformation of the PIM to PSM J2EE Architecture

To ensure an adequate and coherent transformation from the PIM to the PSM some basic rules were set up:

1. Application of all the capabilities of the J2EE server: no server-sided plain old java objects (POJOs), no bean-managed persistence (BMP)
2. Favour declarative over programmatic solutions: use the mightiness of the deployment descriptors
3. Use sustainable and open standards for all interfaces: SOAP/XML for client-server and server-server communication

4.2 Design of the Persistent Domain Model

Following the basic transformation rules the data access to the persistence layer is implemented using container-managed persistence (CMP) thus the container requests, fetches and manages the life-cycle of the entity beans. This reveals performance advantages, bean pooling optimization, and no mix up of framework code and code for the persistence layer.

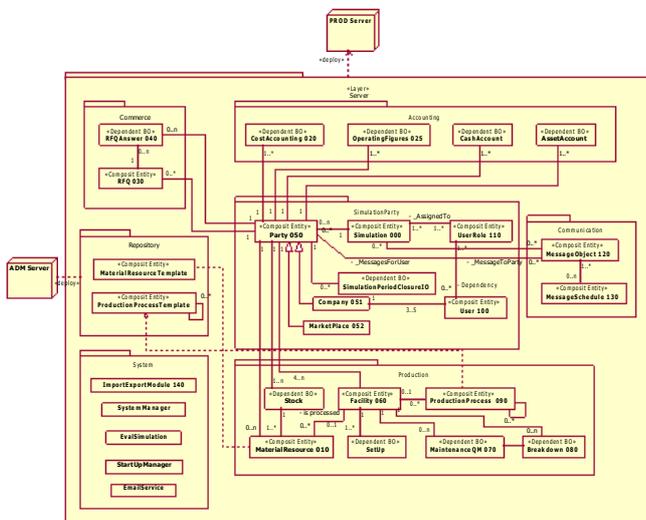


Figure 3: PSM of persistent objects as J2EE design model for implementation (partial view)

The relationship between persistent objects are defined declarative in the deployment descriptor and are managed by the J2EE container: referential integrity, cardinality, cascading delete, and using EJB Query Language (EJB QL) for portable query code [12]. To ensure performance, extensible and scalability, the components and its business (domain) objects are implemented as Composite Entity [2, pp 391] which enables the object-oriented aspect of capsulation of business objects further more. For non-database

persistence like XML feeds, legacy systems, LDAP repositories or flat-files, Data Access Objects [2, pp 460] are implemented.

4.3 Import Facilities using JCA and JMS

Using the J2EE technology platform enables synchronous and asynchronous access via JCA and Java Message Service (JMS) [14] for the import of data and process information of existing e-business or ERP systems. The JMS API enables direct access to external JMS-compliant systems and the implementation of the JMS import facilities manifest as minor expenditure. JCA resource adapters are available for all major systems as middleware infrastructure like IBM MQSeries [10] or ERP systems like SAP's new Netweaver Architecture [30] and can be plugged-in for data import as needed.

6. CONCLUSION

The presented component-based framework for distributed business simulations in e-business environment offers a new conceptual way as an extensible micro-architecture to implement this framework once and extending it to future requirements of diverse domains for simulation and e-business solutions. The pattern driven micro-architecture ensures further development of the framework without pitfalls of scalability or performance problems. We believe new e-learning solutions for preserving and transferring the inherent dynamic system knowledge of rapid changes in e-business will adopt similar strategies as the frameworks introduced in this paper. Future research in combining our simulation framework with Learning Object Metadata for export to other simulation and e-learning systems or for direct access for the persistence layer would widen the framework for even more interoperability.

REFERENCES

- [1] Albores- Barajas, A. et al., "Generic Business Process Models for E-Business in Manufacturing companies: is simulation useful?", *Engineering Management Conference, IEEE*, 8-20 Aug. 2002, Vol.2, pp685-690, 2002.
- [2] Alur, D. et al., "Core J2EE patterns – Best Practice and Design Strategies", Upper Saddle River: Sun Microsystems/Press, Pentrice Hall, 2003.
- [3] Arsanjani, A., "Developing and Integrating Enterprise Components and Service", *Communications of ACM*, Vol. 45, No. 10, pp31-40, 2002.
- [4] CORBA Home, www.corba.org
- [5] Cypser, R.J., "Evolution of an open architecture", *IBM Systems Journal*, Vol. 31, No. 2, pp161-188, 1992.

- [6] Derntl, M., Mangler, J., "Web Services for Blended Learning Patterns", *Proceedings of 4th IEEE International Conference on Advanced learning Technologies (ICALT'04)*, August 30-Sept 1, Joensuu, Finland, 2004.
- [7] Fremantle, P, et al., "Enterprise Services", *Communications of ACM*, Vol. 45, No. 10, pp77-82, 2002.
- [8] Gamma, E. et al., "Design Patterns –Elements of Reusable Object-Oriented Software", Indianapolis: Addison-Wesley, 1995
- [9] Hampel, A., Bernroider, E.W.N, "Strategic Software Engineering Decisions Concerning the System's Architecture and the Development Environment for an Internet-Based Business Simulation", *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Vol. 2004, No. 1, pp. 64-65, 2004.
- [10] IBM, "Patterns for e-business", <http://www-106.ibm.com/developerworks/patterns/u2b/at2-product-w2k.html>
- [11] IEEE , "Draft Standard for Learning Object Metadata", IEEE 1484.12.1-2002 (LOM), 2001.
- [12] J2EE, Java Version 2 Enterprise Editions, <http://java.sun.com/j2ee>
- [13] JCA, Java Connection Architecture, <http://java.sun.com/j2ee/connector>
- [14] JMS, Java Message Service, <http://java.sun.com/products/jms>
- [15] Lewis, G.R., "CORBA 2.0: universal networked objects", *ACM StandardView*, Volume 3, Issue 3, pp102-106, 1995
- [16] Macromedia, Flash MX, <http://www.macromedia.com/software/flash>
- [17] Matyas, K., "Planspiel Produktionssteuerung - Simulation Game in Production Control", PhD Thesis, Austria , Vienna University of Technology, 1989.
- [18] Matyas, K., Hampel, A., "Internet-based Business Simulation for the International Education of Industrial and Systems Engineers – An Interdisciplinary Scientific Long-Term Project", *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Vol. 2003, No. 1, pp. 872-873, 2003.
- [19] Mead, G.C., "Business Simulation – Dynamic Tool or Enjoyable Pastime?", *Journal of Accountant*, No. 121, pp84-85, 1966.
- [20] Nayak, N. et al., "Virtual enterprises: building blocks for dynamic e-business", *Australian Computer Science Communications*, Vol. 23 No. 6, pp80-86.
- [21] Nurmi, S., Lainema, T, "Facilitating the Understanding of Business Processes with Simulation Game", *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Vol. 2003, Issue. 1, pp. 2163-2170, 2003.
- [22] OMG, Model Driven Architecture, <http://www.omg.org/mda>
- [23] Ortiz, A. et al., "Enterprise Integration –Business Process Integrated Management: a proposal for a methodology to develop Enterprise Integration Programs", *Computers in Industry*, 40(1999), pp155-171, 1999.
- [24] Patten, Th.H., "Business Simulation in Industry and University Education", *Personnel Journal*, No. 41, pp566-567, 1962.
- [25] Riis, O. "Games for Production Management", <http://iproduct.auc.dk/x-proj/gamespm/www-games.html>, *International Workshop of the IFIP WG 5.7: Experimental Interactive Learning in Industrial Management.*, 22-24 May 2003, Aalborg, Denmark, 2003.
- [26] RUP, Rational Unified Process, <http://www.rational.com>
- [27] Singh, N., "Unifying Heterogeneous Information Models", *Communications of ACM*, Vol. 41, No.5, pp37-44, 1998.
- [28] SOAP, W3C XML Protocol Working Group, <http://www.w3.org/TR/soap>
- [29] Stecher, P., "Building business and application systems with retail application architecture", *IBM Systems Journal*, Vol. 32. No. 2, pp278-306, 1993.
- [30] Stiefbold, O., Platzer, U., "What's New in SAP's Technologies?", *Javamagazin*, 6/2003, pp45-51, 2004.
- [31] Sutherland, J., Heuvel, J., "Enterprise Application Integration and complex adaptive systems", *Communications of ACM*, Vol. 45, No. 10, pp59-64, 2002.
- [32] Wang, H., Zhang, D, "MDA-based Development of E-Learning System", *Computer Software and Applications Conference, COMPSAC 2003. Proceedings. 27th Annual International* , 3-6 Nov. 2003 , pp684-689, 2003.
- [33] Zachmann, J.A., Sowa, J.F., "Extending an formalizing the framework for information systems architecture", *IBM Systems Journal*, Vol. 31. No. 3, pp590-650, 1992.

A Global Perspective of Business Usage of Web Technology

Chang Liu¹, Wayne Huang²

¹ College of Business, Northern Illinois University, DeKalb, IL 60115, USA, cliu@niu.edu

² College of Business, Ohio University, Athens, OH 45701, USA, huangw@ohio.edu

ABSTRACT

Despite the recent downturn in Internet and telecommunication sectors, e-commerce is forecasted to continue to grow and corporate websites have remained as important communication channels for both businesses and consumers. However, many multi-nation corporations (MNCs) use the same business websites for their business transactions in all of their business locations, even across national boundaries. Does this kind of "one website fits all" mentality of using Web technology really work well in reality? Does national culture matter in terms of influencing the adopting of Web technology for commercial usage in MNCs? Whether corporations in different countries/cultures use Web technology differently in their commercial websites? This paper intends to explore the above important research issues through examining and comparing the usage of commercial websites among corporations in two different countries: Singapore and the United States.

Keywords: Web, Electronic Commerce (Business), USA, Singapore, Global

1. INTRODUCTION

Over the last few years the Internet has evolved to a platform that is enabling a new generation of businesses. Companies have realized the tremendous opportunities the Internet offers and have established an online presence with sound business models. In general, e-commerce can be defined as buying, selling of products, services, and/or information that use the Internet and Web technology (Hof & Hamm, 2002). E-commerce is proved to have abilities to enhance economic growth, reduce barriers of market entry, improve efficiency and effectiveness, keep inventories lean, and cut cost (Madden & Coble-Neal, 2002).

At the same time, consumers have found the power on the Internet: convenience, greater choice, vast amount of information, and time savings, are not going to let challenging economy stop them to taking advantage of it. The number of business to consumer (B2C) spending is expected to reach \$95 billion in 2003, and could reach \$250 billion in 2005. Moreover, the business of business (B2B) online transactions now stand at \$2.6 trillion, not the originally expected \$1.3 trillion, and analysts are revising their estimates of productivity gains sharply upward (Mullaney, Green, Arndt, Hof, 2003). Indeed, e-commerce has transformed entire industries globally and the Net is really "changed everything" and even more startling changes are on the horizon.

2. RESEARCH BACKGROUND

This paper intends to study and compare the Internet's business usage among corporations in the U.S. and Singapore, in order to explore whether national culture could be an important factor influencing the Internet usage in business corporations. These two countries are from two different continents across the globe and can be viewed to best represent members of its respective continent.

The U.S. has the largest number of online users in the world. The U.S. broadband subscriptions which allow speedy connection to the Internet doubled since 2001 and are growing at 56% annually. It is expected to reach 40 million households at the end of 2003 (Forrest Research, 2003). Studies have found that Internet and Web are not only used to generate revenues, but also used as tools to increase productivity in the U.S. In fact, productivity growth is doubled as the Web proliferated since 2000 (Mullaney, Green, Arndt, and Hof, 2003). More and more companies are utilizing the Internet medium for advertising purpose to attract customers online. According to AdRelevance report in 2003, a division of market researcher Nielsen-NetRatings, big advertisers are showing renewed interest in online advertising after the downturn of the economy. Purchases of Web ads rose to \$3.2 billion in the first half of the 2003, jumped from \$2.8 billion in the same period of 2002. "Search-related" advertisements, which appear alongside results on well-trafficked Web-search sites, are particularly popular (Mangalindan, 2003). This is not only helping boost the sales and fortunes for the businesses, it also helps businesses to target their potential customers.

Singapore, on the other hand, is the seventh in terms of preparedness for e-commerce. Singapore government has also been extremely supportive of its e-commerce activities. It has played a multi-faceted role of being both the user and promoter of e-business application in the country. Concrete steps have been taken to position Singapore as a global node and regional hub in e-commerce. The Electronic Commerce Master plan was launched in 1998 and the legal framework has been updated through the passing of the Electronic Transaction Act. The government strives to build up research capability by encouraging leading e-commerce research companies to be based in Singapore. Singapore encourages local establishments to link up with international research centers in order to learn from

others and to share their expertise. The number of Internet dial-up subscribers in Singapore increased from 139,000 in January 1997 to over 1,943,800 in 2001, placing Singapore at the seventh out of 60 countries in terms of preparedness for e-commerce. Statistics have shown positive outbreak of revenues from e-commerce. For instance, according to the Info-communications Authority of Singapore (IDA), the revenues from e-commerce sector was to top about US\$1.4 billion in 2001 as compared to US\$659 million in 2000. Statistics have also shown a higher amount of people buying items such as airline tickets, books and CDs, and reserving hotel rooms through the Web.

3. RESEARCH QUESTIONS

Although commercial websites are expected to grow to 200 millions by 2005 (Palmer, 2002), the business usage of Web technology to reach a large number of customers worldwide remains a big challenge for many multi-nation corporations (MNCs). Customized Web design with multi-language support alone, however, is not enough. Website builders must also consider political, cultural, social, and legal differences in the global environment. For example, culture provides people with a sense of identity and an understanding of acceptable behavior within society. Culture includes values, shared beliefs, or group norms that have been internalized by individuals with some modification (Hofstede, 1980). In addition, culture influences peoples' perceptions and their interpretation of the world. It influences their expectations, values, beliefs, attitudes, and ultimately their behaviors in everyday life (Adler, 1991). Groups of people have unique backgrounds because of their shared history, economy, geography, religion, and demographics. Therefore, it is important for MNCs to understand how culture may influence the business usage of Web technology on a global scale.

The U.S. is traditionally recognized as having a strong culture based on individualism that stresses independence of each individual within a society. On the other hand, Singapore focuses more on collectivism and emphasizes the goals of the group over personal goals, stresses conformity and in-group harmony, and defines the self in relation to the group (Lu, Rose, & Blodgett, 1999; Triandis, 1995).

Although several studies addressed the usability and design of Web sites (e.g., Palmer 2002, Paerrow 2000, Nielsen 2000, Shneidermann 1998), none have extended their studies to across national boundaries to examine the website usages in a global environment to our best knowledge. Therefore, a study in this topic will help fill the gaps. This study intends to examine and compare the Web usage of businesses among corporations in the U.S. and Singapore. The two countries are from two different continents across the globe and can be viewed to best represent members of its respective continent with its

culture background. The specific research questions are:

- *What are the main features of the corporations' web sites in these two countries?*
- *What are the important contents that are normally included in most commercial web sites in these two countries?*
- *Is the usage of web sites in business corporations related to the characteristics of the industries that the corporations belong to?*
- *Are there any differences of using web sites in these two countries?*
- *Does culture play a role in the site differences in these two countries?*

4. RESEARCH METHODOLOGY

In order to do a comparative analysis of commercial website usage among corporations in the two countries, a common assessment framework was used to collect data and conduct the comparative analysis. The framework consists of 14 main categories describing the commercial usage of web technology (e.g., Cheung and Huang, 2002; Liu, Arnett, Capella, and Beatty, 1997), as shown below.

- **Company overview/information:** Brief history; how the company organized and operates; company capability; general business environment; challenges, etc.
- **Products/services:** Main products and services.
- **What's new:** News releases for the products and services; updated news; pertinent industry information; new announcements; hot topics, etc.
- **Search:** Search engine for users to locate available interested items within the company's home page.
- **Employment/Career opportunities:** Available opportunities for certain positions and their descriptions.
- **Interactive feedback:** Requesting viewers to respond about their feelings/suggestions about a company and/or web home page.
- **Customer service/assistance:** Answering questions; sending forms to relevant person for products/services support; providing necessary help, etc.
- **Index/directory:** Company's Web site direction or map; location distribution; department addresses; office phone numbers, etc.
- **Financial/Investment facts:** Detailed financial performance
- **Links to other sites:** Connection with Internet yellow page, weather, stock, quotes, news groups or third party vendors.
- **Online business:** For information transformation on ordering/banking through the company's home page.
- **Guest book:** Requesting customers to sign in by providing their names and addresses.
- **FAQ:** Questions and answers for the frequently asked questions about the company, its products, its services, its website, or something else.

- Messages from CEO: Greetings and brief introductions from CEOs

Then the authors of this paper accessed to more than 700 commercial web sites and collected data based upon the framework. More specifically, Fortune 500 corporations were chosen as the data points representing US corporations. These companies' Web sites generally provide a leadership of using information technologies (Liu et al, 1997). In Singapore, a total of 250 Web sites were randomly selected based on queries results of using search engines Yahoo and AltaVista. These corporations' web sites represented 11 main industries in Singapore (Cheung and Huang,2002). In addition, an extensive search has also been done to collect the most updated information on e-commerce in these two countries, using mainly the search engine of GOOGLE and OCLC library search.

5. RESEARCH FINDINGS

5.1 Company Size with regards to E-commerce embracement

U.S.

Results from the United States regarding e-commerce embracement seem to differ from those of Singapore. Many small sized companies are moving faster and getting dramatic results using direct selling on the Web. They tend to take more advantage of the low cost advertising and saved space on the Web to generate revenues, whereas most of the large companies such as the Fortune 500 tend to mainly announce their presence and promote their companies' image, enhance public relations, etc.

Singapore

On the other hand, the growth of e-commerce has been unevenly distributed. Large enterprises have generally been more ready to embrace e-commerce than small enterprises. 96% of the large enterprises are equipped with Internet access as compared to 86% for the small enterprises. Also, about 61% of the large firms have their own websites, compared to only 32% among the small enterprises.

5.2 Industry's Web Presence

U.S.

As indicated above that the Fortune 500 companies were selected to represent the United States because they usually provide the leadership in the use of information technology. Industry types are classified into 11 divisions by Standard Industry Code (SIC) of which, the Fortune 500 comprises of 9, with statistics of percentage Web presence shown below in Table 1.

In the case of the Fortune 500, the Transportation industry seems to be the leader in Web presence. However, it should be noted that the industry category is

classified differently from how it is for Singapore. For Singapore, the Retail and Wholesale industries are combined into one; whereas for the Fortune 500, the Retail Trade and Wholesale trade are distinguished, thus making the statistics percentage lower than that of Transportation industry. Nevertheless, it is evident that the Retail/Wholesale industry will place the highest on the scale.

Table 1: Industry's Web Presence in U.S.

Industry	WWW presence
Transportation, communication, electric, gas, sanitary services	87.10%
Retail Trade	68.10%
Manufacturing	63.60%
Finance, insurance, real estate	53.30%
Services	51.40%
Agriculture, forestry and fishing	47.40%
Wholesale trade	42.30%
Construction	14.30%
Mining	0.00%

Another internet-related activity big company involved in is e-procurement. Even though the connection on the Internet has been established, most companies do not seem to take full advantage of it for e-procurement. According to a study by Hurwitz Group, less than one-third of Fortune 500 businesses order strategic goods from online suppliers. VP of e-business Strategies at Hurwitz Group said despite the promise of tremendous benefits from e-Procurement, including perceived savings of up to 50 percent in purchasing costs, most companies are not measuring these anticipated benefits.

Most e-procurement systems being utilized lack support for complex procurement processes, including contract management and negotiation. The study from Hurwitz Group stated that companies may work with thousands of suppliers, but fewer than 100 suppliers were linked to their e-procurement system. The study also found most companies that do make purchases online are performing indirect purchases of maintenance, repair, and operational goods and services instead of using e-procurement to purchase direct goods to automate production or manufacturing processes.

Singapore

The SES – Security Exchange of Singapore classified all listed businesses in Singapore into 5 big categories: industrials, finance, hotels, properties, and plantation (Cheung and Huang, 2002). Based on this categorization, another set of categories of businesses has been further developed, dividing the Singapore businesses into 11 categories, including:

1. Retail/Wholesale Trade
2. Information Technology/Computer
3. Manufacturing

4. Public Utilities
5. Tourism
6. Transportation
7. Finance/Banking
8. Design/Advertising/Press/Media
9. Properties and Holding (Real Estate)
10. Human Resource/Consultation
11. Other

Retail/Wholesale industry takes the lead in WWW business presence with 27.6%. This is an interesting finding for Singapore since the industry type that takes the lead in Web presence is not Information Technology, which places 2nd on the scale with 22.8% Web presence. This difference is explained by the fierce competition in retail/wholesale industry. Singapore is a small country with the population of about 4 million (The World Fact Book 2002), with many big international retail corporations competing against each other. With such tight competition, small and medium sized companies have turned to e-business as their solution since the Internet provides inexpensive and mass marketing channels, versus the unbearable cost of advertising on TV as well as other mass media. Another important factor to note is the Singapore government has been very supportive of the development of its IT industry. The IT industry has received substantial government funding as well as preferential treatment under government tax and financial regulations and policies (Cheung and Huang, 2002). However, the difference in the total number of companies going online has indicated that the adoption of WWW technology is most likely not driven by government backup, but rather by competitive marketing forces. These two industries are classified as early adopters in the WWW presence. The next categories of industry are classified as early majority in terms of adopting WWW for commercial use. Statistics on each industry's Web presence is summarized below in Table 2.

Table 2: Industry's Web Presence in Singapore

Industry	WWW presence
Retail/Wholesale	27.6%
IT	22.8%
Design	8.80%
Manufacturing & tourism	6.80%
Real Estate	6.40%
Consultation service	5.60%
Transportation	5.20%
Financial/Banking	2.8%
Public Utilities	0.8%

The low percentage of Web presence of Finance/Banking industry at the time of the research is explained by the fact that the industry deals with money and volatile information. Its transactions require a high level of security guarantee. Since the Internet is not truly known for being the safest place to store such

information, i.e. there are still problems with the system being hacked into. Thus, this could explain somewhat the reason for the low figure in Web presence of banking industry. A general chart showing the percentages of companies that have web pages in different industries across the two countries is also composed as following in Figure 1.

Figure 1 shows that the U.S. tends to lead WWW presence in Manufacturing and tourism industry, whereas Singapore leads the WWW presence in its Retail/Wholesale sector. More specifically, the top three industries with WWW presence in US are Transportation/Communication/Electronic, Retail Trade and Manufacturing; the top three in Singapore are Retail/Whole Sale, Information Technology, and Design. Therefore, the two countries differed in web presences of industries.

5.3 Features on Commercial Web Sites

U.S.

The main features found on websites of Fortune 500 include: products and services, company overview, feedback, what's new, financial facts, customer service, search, employment opportunities, guest book, Index or directory, online business, links to other sites, CEO messages and frequently asked questions (FAQ). It indicates that most companies go online to introduce their products/services to a larger market target at a lower cost.

Singapore

Based on the study conducted by Huang (2002), 100% of Singapore web sites contain their companies' names, images, and information of their products/services. The next most popular feature is online information exchange (26.8%) and the third in line is customer support (19.6%). For Singapore Web sites, the least likely purpose of the homepages is to achieve sales on the Internet (17.2%). The result indicates that trading activities on line had not fully bloomed in Singapore. Companies set up homepages mainly to publicize the companies and their images, rather than to take advantage of the profit-enhancing activities available on the Web. The Internet security issues and customer preference of old purchasing behaviors could be the reasons contributing to the lack of interest from most corporations in Internet sales.

Web in Singapore aims more at customer support and information exchange, whereas Internet sales are most emphasized by the Retail industry. There are a number of barriers that create difficulty for the consumer online retail business. One of the barriers is the fact that Asia does not have the same big base of installed personal computers. On average, there are 0.01 personal computers per person versus 0.40 in the U.S., according to IDC. Another barrier is the built-in habit of the users. There are shopping malls everywhere in Singapore and

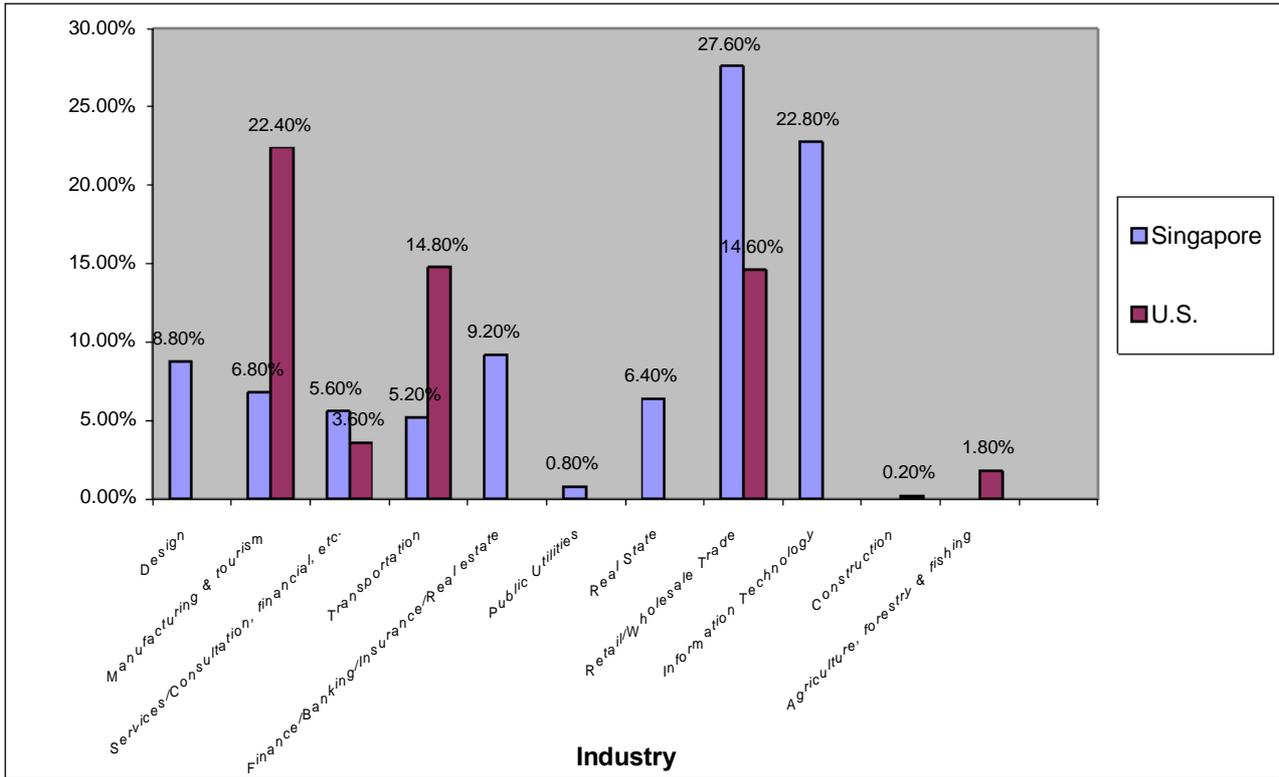


Figure 1: WWW Presence in Singapore and the U.S.

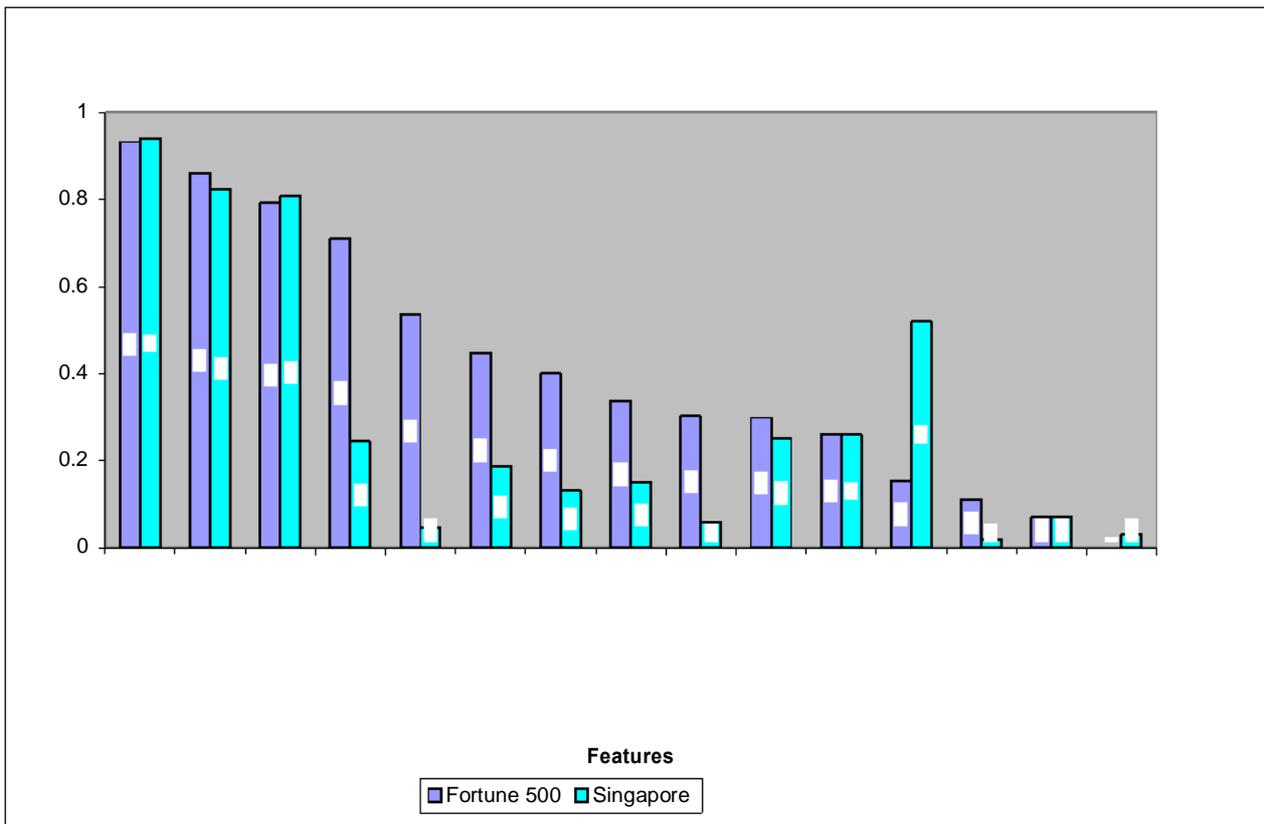


Figure 2: Web Features in Singapore and the U.S.

consumers generally like to feel and touch the products, try and test them out and compare the prices prior to purchasing them. Furthermore, a big concern is the use

of credit card online. Even though more people are spending time online, when it comes to purchasing, they are still reluctant due to concerns over the security of

credit card purchases and personal information. The problem with using credit card online has been addressed by a verification service however, called "Verified by Visa", where customers use personalized password to serve as an additional verification point. When they buy something from a website that is "Verified by Visa", they are given access by entering their password as at ATM machines.

Figure 2 provides a comparison of the frequency of features found on Web sites of the two different countries. On average, the top 3 features utilized on any Web site are products/services, company overview, and feedback, which reinforces the main purpose for corporations going online at this point, which is to reach a large market to promote acknowledgement of their products/services, and to pronounce the existence of the company. Feedback also allows the companies to get customers' ideas and opinions on how they are doing and what could be improved for them to better serve their customers. Other features are not uniform across the two countries. In other words, they are on websites of one country but not found as common on websites of another country. Customer service is also present on all websites but appears to be the lowest in Singapore, which indicates the fact that Singapore companies need to redesign their web site in such a manner that can provide customer services to the extent possible.

6. DISCUSSIONS AND IMPLICATIONS

Our cross-country comparison study indicated that on the usage of commercial web sites of corporations in the two different countries is different from each other, in terms of the three specific aspects, web presence in different industries, web features used in their web sites, and the relationship between corporation size and commercial web sites, etc. Corporations have spent millions of dollars in developing web sites for commercial purposes in the past years. Different ways of using commercial web sites would reflect different needs and characteristics of local businesses in the two countries. As a result, the one-size-fits-all method of designing and using commercial web sites may not work well when crossing national geographical boundaries and/or cultures. Therefore, when designing and developing commercial web sites and/or e-commerce systems for corporations in different countries, national culture could be an important factor to be considered in order to design a successful and effective commercial web site for corporations in another culture/country.

The majority of companies in Singapore and the Fortune 500 companies are using the Internet medium mainly for advertising purpose, to announce their existence rather than to generate revenue. However, they have all realized the huge potential provided by their Web

presence and have implemented strategies to attract more hits to their sites, as well as to generate more revenues from their Web presence. As far as the features on the Web sites, even though the companies are physically dispersed, their Web presences share fairly similar standards for content display. The features most frequently found on all websites across the two countries are Products/Services, Company Overview, Feedback and Customer Service. The percentage of their use may differ, i.e. Singapore has a low percentage on Customer Service compared to the Fortune 500; however, in general, they all mean to expose the companies to the public to promote knowledge of the public toward their products/services provided, as well as to enhance public relations, to reach global customers, and to get feedback from customers to improve the way they can do business and better serve the customers. Regardless of the fact that industries are classified somewhat differently across the two countries; however, the Retail industry seems to be the one with the highest Web presence in all, also the one that generates the most revenue from its web presence compared to other industries. Despite any drawback in the economy or concern over security on making purchases online, it is predicted that revenues generated by businesses conducted via the Web will be on the rise across the two countries, defining the new age of technology.

REFERENCES

- [1] Adler, N. J. (1991), *International dimensions of organizational behavior* (2nd Ed.), Boston: PWS-Kent.
- [2] Cheung, W. M. and Huang, W. (2002). "An Investigation of Commercial Usage of the World Wide Web: A Picture from Singapore", *International Journal of Information Management*, pp. 2-18.
- [3] Frost, E. and Sullivan, R. (2002), "Singapore's e-biz market poised for growth", <http://asia.internet.com>
- [4] Hof, R.D. and Hamm, S. (2002). "How e-biz rose, fell, and will rise anew", *BusinessWeek*, May 13, pp. 64-72.
- [5] Hofstede, G. (1980), *Culture's Consequences: International differences in work-related values*, Sage, Beverly Hills, Calif.
- [6] Huang, W. (2002). "What web features and functions are used by Australian corporations in their web sites? – a corporate framework and an empirical investigation",
- [7] Liu, C., Arnett, K.P., Capella, L.M., and Beatty, R.C. (1997). "Web sites of the Fortune 500 Companies: Facing customers through homepages", *Information & Management* 31(1), pp. 335-345.

Due to the limited space, readers may send a request to the authors for other references.

A Personalized Recommender System Based on Explanation Facilities Using Collaborative Filtering

Do Hyun Ahn, Hee ae Lee

School of Business Administration, Kyunghee University Seoul, 130-701, Korea
adh@khu.ac.kr, leeheae@hanmail.net

ABSTRACT

Collaborative filtering (CF) is the most successful recommendation method, but its widespread use has exposed some limitations, such as sparsity, scalability, and black box. Many researchers have focused on sparsity and scalability problem but a little has tried to solve the black box problem. Most CF recommender systems are black boxes, providing no transparency into the working of the recommendation. This research suggests an improved CF recommender system with explanation facilities to overcome the black box problem. Explanation facilities make it possible to expose the reasoning and data behind a recommendation. Therefore, explanations provide us with a mechanism for handling errors that come with a recommendation. Furthermore, it is proposed to use web usage mining and product taxonomy to enhance the recommendation quality for e-commerce environment. For such purposes, it is developed a recommender system named WebCF-Exp, Web usage mining driven Collaborative Filtering with Explanation facilities. To test the performance of WebCF-Exp, EBIB research internet shopping mall and explanation interfaces are developed. Experiments are conducted with the data provided by EBIB Research Internet shopping mall.

Keywords: Collaborative Filtering, Recommendation system, Explanation, Internet shopping mall

1. INTRODUCTION

The continuous growth of the Internet and e-commerce has allowed companies to provide customers with more choices on products. Increasing choice has also caused product overload where the customer is no longer able to effectively choose the products he/she is exposed to. A promising technology to overcome the product overload problem is recommender systems that help customers find the products they would like to purchase. To date, a variety of recommendation techniques have been developed. Collaborative Filtering (CF) is the most successful recommendation technique, which has been used in a number of different applications such as recommending movies, articles, products, Web pages, etc. [1,2,3,7,10,11]. However, its widespread use has exposed some limitations, such as sparsity, scalability, and black box. Many researchers have focused on sparsity and scalability problem but a little has tried to solve the black box problem. According to Herlocker et al.[6], most CF recommender systems are black boxes, providing no transparency into the working of the recommendation.

This research suggests an improved CF recommender system with explanation facilities to overcome the black box problem. Explanation facilities make it possible to expose the reasoning and data behind a recommendation. Therefore, explanations provide us with a mechanism for handling errors that come with a recommendation. Furthermore, it is proposed to use web usage mining and product taxonomy to enhance the recommendation quality for e-commerce environment. Web usage mining populates the rating database by tracking customers' shopping behaviors in the Web, thereby leading to better

quality recommendations. The product taxonomy is used to improve the performance of searching for

nearest neighbors through dimensionality reduction of the rating database. For such purposes, it is developed a recommender system named WebCF-Exp(Web usage mining driven Collaborative Filtering with Explanation facilities). To test the performance of WebCF-Exp, EBIB (e-Business & Intelligence Business) research internet shopping mall and explanation interfaces are developed. Experiments are conducted with the data provided by EBIB Research Internet shopping mall.

The remainder of this paper is organized as follows. Section 2 reviews the past researches related to recommender systems and explanation. Section 3 and section 4 provide our research framework and system architecture. Section 5 describes experimental evaluation and section 6 finally provides some conclusions and future works.

2. LITERATURE REVIEW

2.1 CF-based Recommender Systems

CF-based recommender systems recommend products to a target customer according to the following steps [11]: (1) A customer provides the system with preference ratings of products that may be used to build a *customer profile* of his or her likes and dislikes. (2) Then, these systems apply statistical techniques or machine learning techniques to find a set of customers, known as *neighbors*, which in the past have exhibited similar behavior (i.e. they either rated similarly or purchased similar set of products). Usually, a neighborhood is formed by the degree of similarity

between the customers. (3) Once a neighborhood of similar customers is formed, these systems predict whether the target customer will like a particular product by calculating a weighted composite of the neighbor's ratings of that product (*prediction problem*), or generate a set of products that the target customer is most likely to purchase by analyzing the products the neighbors purchased (*top-N recommendation problem*). These systems, also known as *the nearest neighbor CF-based recommender systems* [2,10,11] have been widely used in practice. However, its widespread use has exposed some limitations, such as sparsity, scalability, and black box.

2.2 Web Usage Mining

Web usage mining is the process of applying data mining techniques to the discovery of behavior patterns based on Web log data for various applications. The overall process of Web usage mining is generally divided into two main tasks: data preprocessing and pattern discovery. Mining behavior patterns from Web log data needs the data preprocessing tasks that include data cleansing, user identification, session identification, and path completion. Mobasher et al.[9] presented a detailed description of data preprocessing methods for mining Web browsing patterns. The pattern discovery tasks involve the discovery of association rules, sequential patterns, usage clusters, page clusters, user classifications or any other pattern discovery method. Lee et al.[8] provided a detailed case study of clickstream analysis from an online retail store. To measure the effectiveness of efforts in merchandising, they analyzed the shopping behavior of customers according to the following four shopping steps: product impression, click-through, basket placement, and purchase. It has been recognized that Web usage mining gave better recommendation quality in the CF recommendation procedures [1,3].

2.3 Product Taxonomy

Product taxonomy is practically represented as a tree that classifies a set of products at a low level into a more general product at a higher level. The leaves of the tree denote the product instances, SKUs (Stock Keeping Units) in retail jargon, and non-leaf nodes denote product classes obtained by combining several nodes at a lower level into one parent node. The root node labeled by All denotes the most general product class. For example, Figure 1 shows product taxonomy for an internet shopping mall, where "SKU00", "SKU09" and "SKU10" are classified into "Outwears", and etc. A number called level can be assigned to each node in the product taxonomy. The level of the root node is zero, and the level of any other node is one plus the level of its parent. Note that a product class at a higher level has a smaller level number. The product taxonomy of Figure 1 has four levels, referred to as level 0 (for root), 1, 2, and 3. Recently, the usage of product taxonomy in data

mining has been emphasized by many researchers [3,5,7] since it reflects domain specific knowledge and may affect the results of the analysis.

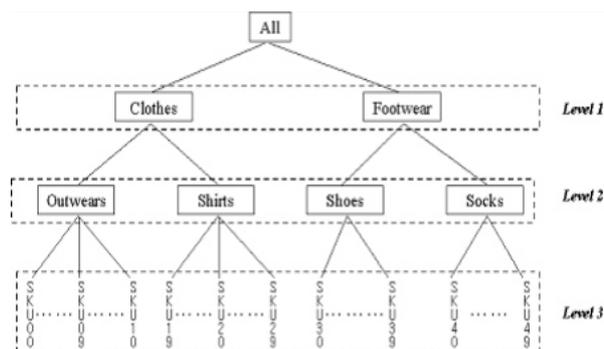


Figure 1. Example of product taxonomy

2.4 Explanation

Most CF recommender systems are black boxes, providing no transparency into the working of the recommendation. Therefore, a user is given no indicators to consult to determine when to trust a recommendation and when to doubt one. The problem has prevented acceptance of CF-based recommender systems in all but the low-risk content domains. Explanation facilities make it possible to expose the reasoning and data behind a recommendation. Therefore, explanations provide us with a mechanism for handling errors that come with a recommendation. Building an explanation facility into a recommender system can benefit the user in many ways. It removes the black box from around the recommender system, and provides transparency [4,6]. According to Herlocker et al.[6], some of the benefits provided by the explanation facilities are:

Justification: User understanding of the reasoning behind a recommendation, so that he/she may decide how much confidence to place in that recommendation.

User Involvement: User involvement in the recommendation process, allowing the user to add his knowledge and inference skills to the complete decision process.

Education: Education of the user as to the processes used in generating a recommendation, so that he/she may better understand the strengths and limitations of the recommendation capability of the system.

Acceptance: Greater acceptance of the recommender system as a decision aide, since its limits and strengths are fully visible and its suggestions are justified.

3. METHODOLOGY

WebCF-Exp is a CF-based recommender methodology based on Web usage mining, product taxonomy, and explanation facilities to improve the recommendation quality and system performance of current CF-based recommender systems. The overall procedure of WebCF-Exp consists of two main methods as shown in Figure 2: recommendation method and explanation

method. The input data consist of Web server log files, product database and purchase database. The endmost output is the personalized product recommendation list with explanation facilities.

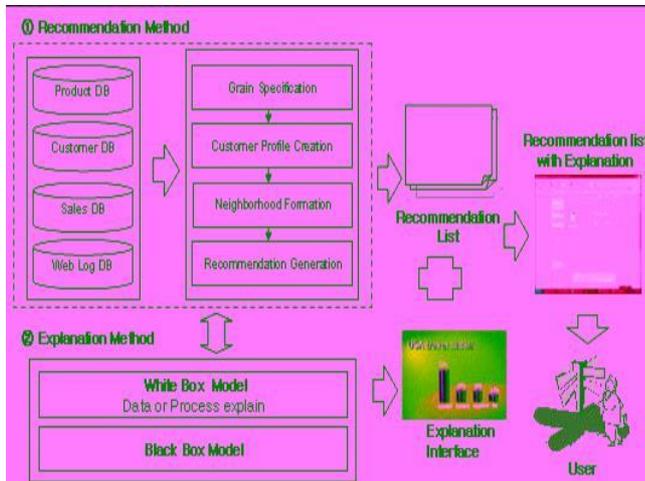


Figure 2. Overall procedure of WebCF-Exp

The recommendation method is divided into four phases: grain specification, customer profile creation, neighborhood formation, and recommendation generation. In the grain specification phase, all products in the database are hierarchically grouped based on the level of aggregation (called *grain*) specified from the marketing manager. Such a product grouping enables the following phases to handle products in the reduced dimensional space. Target customer's preference across products are analyzed and used to make customer profile in the customer profile creation phase. Tracking individual customer's previous shopping behavior in an e-commerce site is used to make preference analysis. The neighborhood formation phase is to form a similarity-based neighborhood between a target customer and a number of like-minded customers. Finally, the recommendation generation phase produces the top-N recommendations based on the shopping behavior of neighbors.

The explanation method consists of white box model and black box model. In white box model, we focus on techniques to justify that the recommendation is indeed performing each of the above phases to the satisfaction of the user and his/her current context. Let us examine each of the steps in more detail, focusing on two components that we need to explain: the process and the data. Often, there is not the opportunity or possibly the desire to convey the white box model. In such cases, the black box model is used to produce recommendations. In black box model, we focus on ways to justify recommendation that are independent of the mechanics.

4. WebCF-EXP RECOMMENDER SYSTEM

For the implementation of the proposed recommender methodology, a recommender system is developed using

agent and data warehousing technologies. WebCF-Exp recommender system consists of four agents as shown in Figure 3: Web log analysis agent, Data transformation agent, Recommendation agent, and Explanation agent. Web log analysis agent manages Web log database through periodic collecting, parsing and analyzing Web server log files such as access logs, referrer logs, agent logs and cookie files. Thus, the users can easily access and analyze it like other operation databases. Data transformation agent creates and manages the data mart that provides data indispensable to accomplish recommendation tasks. Recommendation agent makes a personalized recommendation list for each target customer. Explanation agent provides interfaces which expose the reasoning and data behind a recommendation.

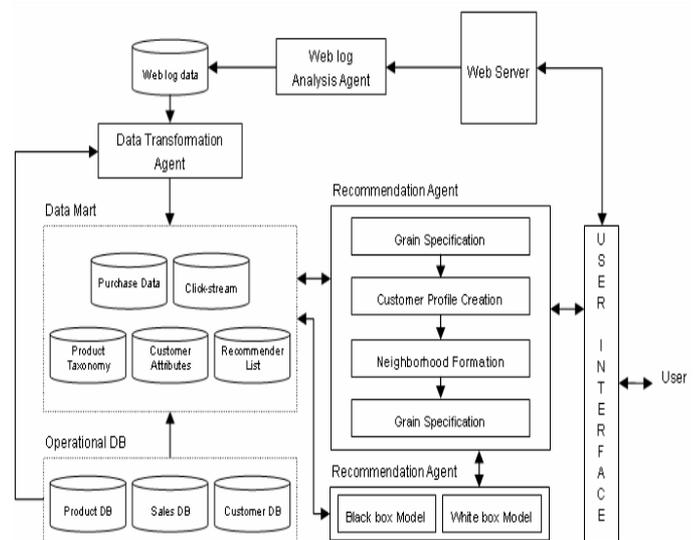


Figure 3. The architecture of WebCF-Exp recommender system

We developed twenty different explanation interfaces to test the recommendation quality. Figure 4 shows one of twenty explanation interfaces.



Figure 4. An example of explanation interface

5. EXPERIMENTAL EVALUATION

In order to evaluate the our suggested recommender

system, we built the prototype of WebCF-Exp recommender system, EBIB Research online shopping mall and explanation interface. We also performed a survey to find the most suitable explanation interface for our recommender system and evaluate the effectiveness of explanation.

5.1 Experimental Design

We conducted the experiment from October 2, 2003 to November 26, 2003, and analyzed the user data between November 18 and November 26, then evaluated the performances of the recommender system. We also surveyed the customers who received our recommendation and saw the explanation interface at the same time. In order to make WebCF-Exp recommender system more understandable, we would explain the process of purchase and recommendation. Programs to perform all our experiments were implemented using Visual Basic 6.0. MS-Access was used to store and process all data necessary for our experiments. EBIB Research online shopping mall was made by MS-SQL Server 2000, WindowsNT2000 Web Server IIS5.0, and ASP(Active Server Page).

Study participants were presented with the following hypothetical situation:

You enter the site of EBIB Research online shopping mall and sign up. You have to submit such data as ID, name, e-mail, birthday, gender and address. When you finish, you will receive 1,000 dollars and can purchase what he want to have. After you log in the site, the server will collect all log data about you. Through the web log agent, the WebCF-Exp recommender system preprocesses web log files and maps click, basket and purchase with "goods_view.asp", "cart_list.asp" and "cart_pay.asp" to extract data related customers' shopping behavior.

Each user was then provided with a different explanation interface, and asked to rate on a scale of 1-5 how likely they would purchase the recommended goods. A successful explanation interface can make a customer more preferable to purchase the recommended goods, and feel the explanation useful and exciting. Therefore we should make it communicate and understand easily and help with the customer's decision making if building an explanation facility into a recommender system. In our research we performed a survey to find the most suitable explanation interface for our recommender system and evaluate the effectiveness of explanation. There are two key research questions that we are interested in answering about the use of explanations with WebCF-Exp recommender system.

The first question is about the understandability and suitability.

Q1: Are the explanation interface understandable and effective?

The goods was explained by the explanation interfaces that we randomly choose from the 20 interfaces, and the question was rated on a 5-point Likert scale ranging from "very difficult" to "very easy."

The second question is about the usefulness of the explanation interface.

Q2: Can the explanation interface increase the acceptance of recommended goods?

The question was rated on a 5-point Likert scale ranging from "not at all" to "very much."

5.2 Experimental Results

(1)Quality comparison of WebCF-Exp and benchmark CF

To compare the quality comparison of WebCF-Exp with that of the benchmark CF algorithm, we performed an experiment to measure *precision* and *recall*. *Precision* and *recall* have been widely used in recommender system research [2,3,11]. Table 1 shows the *precision* and *recall* provided by two algorithms. Looking into the results shown in Table 1, we can see that WebCF-Exp is about 159% and 116% better than benchmark CF, respectively.

Table 1. Quality comparison of WebCF-Exp and benchmark CF algorithm

	WebCF-Exp	Benchmark CF
<i>Precision</i>	0.196	0.123
<i>Recall</i>	0.495	0.425

(2)The result of survey on explanation facilities

The first question is about the understandability and suitability. Table 2 shows mean response and standard deviation of users to explanation interfaces. Looking into the results shown in Table 2, explanation with simple graph and description is better than other explanation types.

The second question is about the usefulness of the explanation interface: Can the explanation interface increase the acceptance of recommended goods? Looking into the results shown in Table 3, the higher customer understands explanation interface, the more helpful customer purchase a product.

Table 2. The result of Q1

	Explanation Interfaces	Mean Response	Std Dev
1	Neighbor ratings (5 stars)	3.806	0.992
2	Discounted item (description)	3.609	1.060
3	The ratio of purchase (graph)	3.536	1.069
4	The ratio of basket (graph)	3.472	1.035
5	The 64% of recommended customers purchased this item	3.443	1.127

	(description)		
6	This item is similar to purchased items in the past (description)	3.417	1.229
7	The best item in the last week (description)	3.277	1.005
8	The ratio of click (graph)	3.261	0.993
9	The ratio of basket (description)	3.247	0.950
10	The preference on Product Taxonomy (graph)	3.217	1.260
11	The ratio of basket (table)	3.210	1.028
12	The ratio of click (table)	3.207	0.942
13	Detailed process (description)	3.173	1.104
14	Recommended by a famous magazine(description)	3.155	1.253
15	The ratio of click (description)	3.152	0.969
16	The preference on Product Taxonomy (table)	3.149	1.051
17	The preference on Product Taxonomy (description)	3.138	1.197
18	The ratio of purchase (description)	3.100	1.150
19	The ratio of purchase (table)	3.049	1.206
20	Recommender system confidence in prediction (description)	3.023	1.116

Table3. The result of Q2

	Explanation Interfaces	Mean Response	Std Dev
1	Neighbor ratings (5 stars)	3.710	1.038
2	Discounted item (description)	3.609	1.060
3	The ratio of purchase (graph)	3.536	1.069
4	The 64% of recommended customers purchased this item (description)	3.371	1.014
5	This item is similar to purchased items in the past (description)	3.277	1.031
6	The ratio of click (graph)	3.261	0.993
7	This item is similar to purchased items in the past (description)	3.253	1.229
8	The ratio of basket (graph)	3.248	1.036
9	The ratio of basket (description)	3.235	0.947
10	Detailed process (description)	3.185	1.108
11	The preference on Product Taxonomy (graph)	3.181	1.251
12	The ratio of basket (table)	3.180	0.989
13	The ratio of click (description)	3.162	0.972
14	The ration of click (table)	3.161	0.926
15	The preference on Product Taxonomy (description)	3.149	1.218
16	The preference on Product Taxonomy (table)	3.138	1.058
17	The ratio of purchase (description)	3.090	1.111
18	Recommender system confidence in prediction (description)	3.058	1.120
19	The ratio of purchase (table)	3.049	1.206
20	Recommended by a famous magazine(description)	2.825	1.250

6. CONCLUSION

Collaborative filtering (CF) is the most successful recommendation method, but its widespread use has exposed some limitations, such as sparsity, scalability, and black box. Many researchers have focused on sparsity and scalability problem but a little has tried to solve the black box problem. Most CF recommender systems are black boxes, providing no transparency into the working of the recommendation. This research suggests an improved CF recommender system with explanation facilities to overcome the black box problem. Explanation facilities make it possible to expose the reasoning and data behind a recommendation. Therefore, explanations provide us with a mechanism for handling errors that come with a recommendation. Furthermore, it is proposed to use web usage mining and product taxonomy to enhance the recommendation quality for e-commerce environment. For such purposes, it is developed a recommender system named WebCF-Exp(Web usage mining driven Collaborative Filtering with Explanation facilities). To test the performance of WebCF-Exp, EBIB (e-Business & Intelligence Business) research internet shopping mall and explanation interfaces are developed. Experiments are conducted with the data provided by EBIB Research Internet shopping mall. Our experiment result shows that WebCF-Exp recommendation system shows better performance than existing CF recommendation systems. And explanation with simple graph and description, showing the evaluation of similar customers, is better than other explanation types. Furthermore, the higher customer understands explanation interface, the more helpful customer purchase a product. Based on these results, an explanation facility added recommendation system is expected to be a useful tool in internet shopping malls.

REFERENCES

- [1] Ahn, D.H., J.K. Kim, Y.H. Cho, "A comparative evaluation of hybrid product recommendation procedures for Web retailers", *APDSI (9th Asia-Pacific Decision Sciences Institute) Conference 2004, Seoul, July 1-4, 2004*.
- [2] Billsus, D., M.J. Pazzani, "Learning collaborative information filters", *In Proc. 15th International Conference on Machine Learning*, pp 46-45, 1998.
- [3] Cho, Y.H., J.K. Kim, "Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce", *Expert Systems with Applications*, Vol.26, pp 233-246, 2004.
- [4] Dahlen, B.J., J.A. Konstan, J.L. Herlocker, N. Good, A. Borchers, J. Riedl, "Jump-starting movielens: User benefits of starting a collaborative filtering system with dead data", *University of Minnesota TR 98-017*, 1998.
- [5] Han, J., Y. Fu, "Mining Multiple-Level Association Rules in Large Databases", *IEEE Transaction on Knowledge and Data Engineering*, Vol.11, No.5, pp 798-804, 1999.

- [6] Herlocker, J.L., J.A. Konstan, J. Riedl, "Explaining Collaborative Filtering Recommendations", *In Proc. 2000 ACM Conference on Computer Supported Cooperative Work*, pp 241-250, 2000.
- [7] Lawrence, R.D., G.S. Almasi, V. Kotlyar, M.S. Viveros, S.S. Duri, "Personalization of supermarket product recommendations", *Data Mining and Knowledge Discovery*, Vol.5, No.1, pp 11-32, 2001.
- [8] Lee, J., M. Podlaseck, E. Schonberg, R. Hoch, "Visualization and analysis of clickstream data of online stores for understanding web merchandising", *Data Mining and Knowledge Discovery*, Vol.5, No.1-2, pp 59-84, 2001.
- [9] Mobasher, B., R. Cooley, J. Srivastava, "Automatic Personalization based on Web usage mining", *Communications of the ACM*, Vol.43, No.8, pp 142-151, 2000.
- [10] Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", *In Proc. Computer Supported Cooperative Work, Chapel Hill, NC*, pp 175-186, 1994.
- [11] Sarwar, B., G. Karypis, J. Konstan, J. Riedl, "Analysis of recommendation algorithms for e-commerce", *In Proc. ACM E-Commerce*, pp 158-167, 2000.

A Process Model for Assessing the Distribution Options for Horizontal E-Business Portals

Tapati Bandopadhyay, Pradeep Kumar, Tapas Mahapatra

Faculty Member (IT & Systems)
ICFAI Business School (ICFAI University), Gurgaon, 122016, India
{tapati, pkgarg, tapas}@ibsdcl.org

ABSTRACT

E-business portal acts as an interface between the e-suppliers and e-customers and many different types of distribution channels are defined individually by various enterprises. The logistics management and decision parameters for distribution models depend on the type of portal e.g. horizontal or vertical. In this paper the focus is on the distribution channels defined for horizontal portals, which are critical aspects of e-business but are not as explored as the other aspects. In this paper, various aspects of e-business models have been analyzed and research reveals that distribution issues need to be looked at with a fresh approach, because the traditional methods do not take into account some typical characteristics of e-business like the range of goods, transactional values and volumes the horizontal portals nowadays handle. Towards this end, three types of distribution channels for a generic horizontal portal have been identified. A decision parameter table has been formulated and used to assess various options for the distribution logistics for the horizontal portals. Also each of these broad categories has been dealt with individually highlighting their salient features along with the advantages and disadvantages associated. The primary objectives for assessment are cost savings and profit maximization of the portal. An assessment process model has been developed on the basis of some key e-business tangible parameters like transactional value, business volume etc. which can be further extended to include specific e-business model dimensions. The parameters are also analyzed subsequently in terms of their availability, size etc. so that the implementation considerations can be realistically made. Finally, the application potential, extendibility and usability of the process model have been explained and it has been shown that this generic model is simple, flexible and specific implementations can lead to e-business portals functioning with a better competitive advantage.

Keywords: Horizontal portals, e-business, distribution models, decision parameters, process model

1. INTRODUCTION

E-business: nowadays complete concept in its own, covers all aspects of a traditional business, including CRM. E-business models, like any traditional business model can be primarily conceived as a revenue-generation model, supported by the other necessary business activities like finance, marketing models etc. ([1], [8]) and it also reflects the company's position in the value chain. The models range from the simplest where a company produces a good or service and sells it to customers, to models more intricately woven, e.g. e-auctioning, brokerage, dynamic pricing or community service-based models where the value chain is not directly or obviously understood or are multi-level hybrids of other simple models.

Whatever the business models [9] are, there is a lot of research evidence about the basic fact that the web has incorporated radical paradigm shifts in the traditional business models [14] and also has given rise to new ones. The web has actually complemented or reinvented tried-and-true models, and has opened hitherto unseen revenue-earning models. [13]

1.1 Various Aspects Of E-Business Models

E-business models have been defined, analyzed and categorized in many different ways and they continue to evolve. When analyzed in the context of references [1], [3] and [14], they essentially include the following aspects:

1. supply chain management
2. customer satisfaction and CRM
3. marketing
4. cost vs. quality of service

Most of the research work on the paradigm shift in e-business models is concerned about the supply chain management, customer satisfaction and CRM, and marketing ([5], [12]). The cost to quality of service aspect is one area which includes procurement and distribution logistics. Interestingly, even though traditional business models give a lot of emphasis onto these elementary aspects, research reveals that many successful e-business companies or portals are still sticking to the traditional models so far as procurement and distribution are concerned. Procurement has some research leverages as research efforts on B2B portals are coming up with interesting concepts like collaborative commerce and ERP II etc. But distribution options are still the traditional ones, which are not letting an e-business company to fully exploit its'unique e-enabled infrastructure. Profitability in E-supply chains (ref. [4], [10]) includes considerations on

confidence of customer on supplier, more availability than possible in the physical world and data integrity. The first two elements are primarily dependent on the value-chain including procurement and distribution issues. This paper handles these issues relating to the distribution models of e-business portals focusing on speed of delivery and customer satisfaction which reflects the quality of service aspect, customer relationship and cost associated with distribution channels. The distribution problem for horizontal portals has been handled here and a solution process model is suggested.

2. PORTALS: HORIZONTAL OR VERTICAL

Horizontal portals have broad, general user base e.g. Indiatimes.com, yahoo.com etc. which are often B2C type in nature. Conversely, vertical portals serve a particular industry or a user community's supply chain, e.g. indiamart.com- the cyber-marketplace for automobile industry. For vertical portals, supply chains are mainly B2B type, quite well-defined by traditional business and can be followed in e-business as well using EDI and on bulk business. But, horizontal portals operate primarily as e-shopping malls wherein a diverse range of products of various companies are available [15]. These portals acting mostly as virtual intermediaries do not prefer intermediate physical warehouses thereby eliminating the necessity of an effective inventory management and control system. But this way, the horizontal portals have to cater to a huge range of demands of individual customers with orders of varied sizes and transactional values, raising a unique distribution problem in itself [6]. While minimizing on the cost of distribution, the customers also have to be satisfied with the speed of order fulfillment and delivery of goods [7].

The traditional mathematical foundation of distribution logistics has been operations research techniques like transportation and assignment models. But the typical characteristics of horizontal portals e.g. the huge range of transactional values, huge variations in the volume of products etc. are neither taken into account nor analyzed in these traditional methods. An attempt is made in this paper to address this particular issue. These aspects have been included in the analysis and are reflected in the set of decision parameters chosen which is elaborated in section 5. Selection amongst various available options to decide on distribution logistics is inherently dependent on the type of business models the portal uses. Following section gives a brief overview of them.

3. HORIZONTAL PORTAL MODELS

Some of the successful business models of the horizontal portals ([1], [8], [11]) are discussed below.

3.1 Direct Producer-Consumer Model

Herein, E-business removes intermediaries from the supply chain, creating a direct, efficient link between producers and consumers. A manufacturer sells directly to customers, increasing profitability while reducing consumer costs by eliminating warehouse and reseller markups. Even the primary contact for service and support also moves through online channels resulting in reduced overhead and speedy service response. Example includes lgezybuy.com [16] - a popular horizontal portal with a big range of white goods manufactured exclusively by LG. Biggest drawback is that the choices of consumers are restricted to just one manufacturer.

3.2 Intermediary Model

Herein, E-business introduces an intermediary for creating an E-Market- the biggest area of e-business with a great number of e-shopping malls and horizontal portals operating. Advantages include availability of options, best and competitive prices, elimination of mid-warehouses and reduction in extra costs like freight etc.

3.3 Customizable Models

They create investing and merchandising opportunities that were previously not available to ordinary customers. Example: baazee.com

Apart from these common models, E-business operates with various other business options too, e.g. brand-building on an e-shopping mall and integrating market segments and product segments. These businesses may or may not deliver the best possible price, but both suppliers and customers find the convenience of a single point of contact worth the exchange.

For optimal functioning of order fulfillment cycle, it is very crucial for these portals to define suitable distribution models for the products/ services they sell. If the portals' business is entirely IT-enabled and service-oriented like the airline/ railways reservation systems or accommodation booking systems etc., the supply chain is not a major issue. For example an on-line train ticket booking portal, which delivers the e-booked tickets overnight, can use any eminent courier company and take the delivery charges i.e. courier charges from the customer, which he/ she is ready to pay for the convenience of the service. But as regards the portals which sell a diverse range of goods from different manufacturers and importers, distribution becomes a crucial bottleneck in the order fulfillment cycle. In order to tackle this, there are a number of options available for the distribution models applicable which are explained in the next section.

4. HORIZONTAL PORTALS: DISTRIBUTION

For portals like yahoo.com selling a huge range of goods from different manufacturers and importers, there are several distribution models available to complete the order fulfillment cycles with a defined efficiency and in a cost-effective way. These models, as perceived by the authors, are mentioned and analyzed below

Model 1) Customer -> Portal -> Manufacturer/importer -> manufacture/ importers' warehouse -> freight carriers(single company goods -> customer

Advantages:

- It eliminates mid-warehouses completely.
- Best quality product is available since company inspected goods are sent straight to customers without getting stored by dealers.

Disadvantages:

- It may not prove to be cost-effective
- Decision parameters have to reflect
- Relative locations of company warehouses and customer.
 - Scale of business.
 - Transactional value.
 - Product price/profitability i.e. .if the profit involved can justify the freight charges or not.

Model 2) Customer -> Portal -> Manufacturer/importer -> manufacture/ importers' warehouse-> exclusive dealers'warehouse -> freight carriers(single company goods) -> customer

Advantages:

- Relative location advantages can be taken.
- Bulk handling will result in lesser freight.

Disadvantages:

- Dealers'commission can not be eliminated. So the company can not get maximum profit and so is unable to pass on cost savings to customers.
- Quality control can not be ensured.

Model 3) Customer -> Portal -> Manufacturer/importer of ordered product ->dealers handling multiple company goods-> freight carriers(multiple company goods) -> customer

Advantages:

- Shared distribution channels / transportation infrastructure/ shared costs is possible between companies.

Disadvantages:

- It can not eliminate dealers'margin.

5. ASSESSMENT OF MODEL OPTIONS

Key business parameters associated with decision making for choosing a model out of the three alternative model options are identified herein as:

T1: Relative location of individual customers and company/ warehouses X transport cost per unit distance

T2: Freight charges: based on weight or volume

T3: Transactional value

T4: Cumulative historical transactional values with the customer (can be procured from historical databases/ marketing data marts)

The business parameters reflect the aspects which have been discussed in the introduction, i.e. the first two parameters *T1* and *T2* include the cost aspect, *T3* and *T4* deals with customer value which in turn reflects indirectly customer satisfaction and relationship parameters. Many other relevant parameters may be included which can be business/ context-specific and the table can be further extended.

5.1 The Process Model

The assessment and decision making process is hereby developed and discussed as a process model which can later on be mapped onto an algorithm or a flow chart with more specific application orientation. In this paper the process model is developed and presented so that a generic understanding of the process can be achieved.

Process model-

Assumptions:

1. Weightings to parameters are to be given in the range of 0-1, i.e. $W1 + W2 + W3 + W4 = 1$
2. All parameters are to be quantified in the same unit e.g. \$

Steps:

1. Calculate the Tangible Option values as follows:

$Value1 = (W1 * T1 \text{ value} + W2 * T2 \text{ value})$

$Value2 = (W3 * T3 \text{ value} + W4 * T4 \text{ value})$

2. Then

Select the best model option with (minimum of *value1* and maximum of *value2*)

3. If the Minimum ($W1 * T1 \text{ value} + W2 * T2 \text{ value}$) and maximum ($W3 * T3 \text{ value} + W4 * T4 \text{ value}$) do not coincide with the same option, then calculate the relative weight of *T1*, *T2*, *T3* and *T4*, i.e. $(W1 + W2) / (W3 + W4)$

4. if $((W1 + W2) / (W3 + W4)) > 1$, then choose the option with Minimum($W1 * T1 \text{ value} + W2 * T2 \text{ value}$), else choose the option with maximum ($W3 * T3 \text{ value} + W4 * T4 \text{ value}$).

6. ANALYSIS OF THE PARAMETERS

The parameters given in the assessment process are analyzed here individually in terms of the following analysis dimensions:

1. Availability of data
2. Computational complexity
3. Volume (data element size)

Table: Parameter analysis matrix

Parameters Vs Analysis dimensions	T1:	T2:	T3	T4:
Availability of data	high	high	high	High
Computational complexity	Low	Low	Low	high
Volume (data element size)	Low	Low	Low	low

From the above analysis, one important resource-related observation is made that is a data warehouse/data mart (for T4) is needed for recording transactional histories of individual customers and calculating the cumulative transactions, which reflects the level of customer loyalty and indirectly customer satisfaction also. These parameters can also be very useful for customer relationship management.

7. HIGHLIGHTS OF THE PROCESS MODEL

The process model analysis shows the following points:

1. A fresh approach, unbiased on the side of traditional operations research techniques is reflected in the process model which takes care of typical e-business aspects discussed in section 1.
2. This process model is flexible, can be combined with traditional logic and a hybrid solution can be developed
3. This model does not want to replace the well-researched areas of transportation and assignment problems but want to complement them by adding typical aspects of e-business.
4. This process model does not necessarily give the most optimal or cost-effective solution, but it gives a realistic and easily usable way to evaluate various options.
5. The process is simple and easy to comprehend.
6. The process can easily be converted to a computable program without any hidden conflicts. The computational requirements are also minimal and therefore the program executing this process will be highly computationally feasible i.e. processing requirements will be less.
7. It can be used by e-business organizations of any size and capabilities.

Some aspects of this process can be extended further, i.e.

1. Inclusion of other parameters exclusively applicable to any particular e-business scenario/ architecture/ company-specific requirements
2. This process only takes into account the tangible values. Intangible ones have not been included because it is a very complex process to convert them into monetary values which is not of the prevalent assumptions made in this process

8. CONCLUSION

The process model proposed in this paper has a realistic application potential, primarily because it is simple, understandable by any business person without any formal mathematical training i.e. knowledge of operations research techniques etc. It is also simple, therefore easy to implement and execute. Even though, it does not guarantee a perfectly optimal solution, it does address the distribution issues typical to various horizontal portals, issues that have so far been rather overlooked. In fact, vertical or B2B portals also can use this model to redefine their distribution channels, which can be more structured and standardized than horizontal ones. The extension possibilities are great as the model is simple, flexible, scalable and can be modified easily to suit specific business/ context requirements.

Coming out of the traditional mindset in dealing with this type of issues, and embracing new ideas and innovations generated in any of the basic infrastructural aspects would be capable of making e-business environment all the more competitive. Such efforts will also allow the new business models exploit their newness in all possible ways by considering all its' aspects with fresh approaches.

REFERENCES

- [1] Afuah, A., C. Tucci, *Internet Business Models and Strategies*, Boston: McGraw- Hill, Irwin., 2001.
- [2] Alt, R., H. D. Zimmerman, "Introduction to Special Section: Business Models.", *Electronic Markets*, Vol. 11, No.1, pp3-9. 2001.
- [3] Boulton, R., T. Elliott, B. Libert & S. Samek., "A business Model for the New Economy." *Journal of Business Strategy*. July-August 2000, pp29-35. 2000.
- [4] Bouwman, H. & Wijngaert, L. van de, "E-business B2C research in context: policy capturing, channel choice and customer value.", Paper presented to *16th Bled eCommerce Conference, eTransformation*, Bled, Slovenia, June 9 - 11, 2003.
- [5] Chen, Z. & A. Dubinsky, "A conceptual model of Perceived Customer Value in Ecommerce: A Preliminary Investigation." *Psychology & Marketing*, Vol. 20, No. 4, pp323-347. 2003.
- [6] Deitel, H. M., P.J. Deitel & K. Steinbuhler, *E-Business & e-business for Managers*. Prentice Hall, 2001.
- [7] Galbreath, J., "Twenty-first century management rules: the management of relationships as intangible assets.", *Management Decision*. Vol. 40, No. 2, pp116-126. 2002.
- [8] Gordijn, J., J. Akkermans, J. van Vliet, "Designing and Evaluating E-business Models", *IEEE Intelligent Systems*, July/August 2001, Vol. 16, No. 4, pp11-17. 2001.
- [9] Hedman, J. & T. Kalling., "The business model concept: theoretical underpinnings and

empirical illustrations.”, *European Journal of Information Systems* Vol. 12, pp49-59. 2003.

[10] Koushik, S. & P. Joodi., “E-business Architecture Design Issues.”*IT Pro* May/June, pp38-43. 2000.

[11] Mahadevan, B., “Business models for internet- Based E-business.”*California Management Review*. Vol. 42, No.4, pp55-69. 2000.

[12] Muther, A., *Customer Relationship Management –Electronic Customer Care in the New Economy*. Heidelberg, Springer. 2002.

[13] Porter, M. E., “Strategy and the Internet.”, *Harvard Business Review*, Vol. 79, pp62-79. 2001.

[14] Rayport, J.F., Jaworski, B.J., *e-business*, New York: McGraw-Hill/Irwin, 2001

[15] Westland, J.C. and Au, G., “A comparison of shopping experience across three competing digital retailing interfaces”, *International Journal of Electronic Commerce*, Vol. 2 No. 2, pp57-69. 1998.

[16] <http://www.lgeazybuy.com/>, accessed as on 11-Jul-2004

[17] [http:// yahoo.shopping.com/](http://yahoo.shopping.com/), accessed as on 25-Jul-2004

An Agile Protocol for E-Commerce

Yong Wang^{1,2}, Qianxing Xiong²

¹ Management School, Wuhan University of Science and Technology, Wuhan 430081, China

² School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430063, China
Witsbank@sohu.com

ABSTRACT

With merchant information increasing rapidly, customers have to spend a lot of time and energies to search for goods they need, as a result, they maybe find nothing. If merchant can respond customer's requirements agilely when the event defined by customer appears, blindness of customer can be avoided, merchant wins business opportunities too. Protocol is the technology fundament for e-commerce. An idea for agile mechanism is introduced, this paper designs an e-commerce protocol with agility, and it discusses its security. This protocol improves e-commerce efficiency and solves activity problems.

Keywords: e-commerce, protocol, agility, E-P executing model

1. INTRODUCTION

Higher efficiency and lower cost of e-commerce makes more and more companies take part in this business mode. Without e-commerce, enterprise can not participate in global challenge on business, it not only changes enterprise mode of product, operation and management, but also influences economic operating of society entirely^[1]. However, development of e-commerce is influenced by informative fundament, security payment, credit system, delivery link and law environment. When customers roam in vast Internet sea to search goods they need, they have to spend a lot of time and energy, as a result, they maybe find nothing. Much overflow information makes e-commerce become lower efficiency and not meet rapidly enterprise application. How to reverse the positive situation to make e-commerce serve enterprise agilely?

Protocol is support technology to guarantee e-commerce security, transaction and application. Many protocols can be concluded three varieties; they are security protocol, transaction protocol and special application protocol. Summarizing these protocols, each does as he likes. Netbill protocol only is used for digital goods sales. SET has more than 3000 lines grammar definition, 28 steps transaction, each step has 6 times RSA. Digicash pays by means of digital cash, because of its uniqueness, it has the same anonymous and non-traceable features as digital cash. The core of Digicash is blind signature^[2]. TLS (Transport Layer Security), IPSEC and PPTP, variation of SSL, are transmitting control protocols over TCP/IP^[3]. Either these protocols are too easy and only can meet some special application, or they are too complicate and lead to lower efficiency and a large expense. They are considered insufficiently in customer requirement, perfection and intelligence, such as lacking credit, negotiation, active service and multi-medium payment.

Protocol is not designed perfectly so that customers

undergo much suffering of overflow information. This paper researches on aspect of business agility; it introduces agile mechanism and makes protocol have agile property.

2. AGILE E-COMMERCE SYSTEM

2.1 Conception on Agility

Let's see an instance for e-commerce, when a customer accesses a merchant web site to buy something, being out of stock, it can not meets customer this time, when the customer comes here next time, it is maybe short yet, but between the two times, perhaps the commodities are in there and sold out again. In this case, it is possible that customer loses the chance not to obtain goods he needs in the end, as a result, customer waste a lot of time, merchant loses business chance too. If merchant can inform customer actively in replenishing stock, this problem can be solved better. This case demands e-commerce system have an ability to serve customer actively, the system is agile. Agility means it executes transaction automatically to finish business activity in coming definition event.

2.2 E-P Executing Model

Following gives E-P executing model for agility mechanism.

Event Definition : Events causing agile service consist of different kind of events.

$$E = \{p_t(o) \in (E_t, E_e, E_c, E_u) | o \in D_T \wedge p \in p(o)\}$$

Where E stands for event set relating to transaction t , $p_t(o)$ stands for the event which transaction t executes operating p on object o , D_t is data set of t , $p(o)$ is operating set of o , E_t, E_e, E_c, E_u respectively stands for time event, external event, complex event and user-defined event.

Transaction Definition : Transaction set deposits a series triggered transaction, which corresponded with different event.

$$T = \{p(o) | \exists o \exists p(o) \in (T_s, T_c, T_m \dots)\}$$

Where T_s , T_c and T_m respectively represent system transaction, customer transaction and merchant transaction.

Trigger Definition : Trigger is a dual series of condition and triggered transaction.

$$R = \{Cond, T\}$$

Where $Cond$ and T respectively stand for condition and triggered transaction.

Agility Mechanism : Agility mechanism is a mate of event and process.

$$A = \{ \langle e, p \rangle | e \in E, p \in P, P = \langle CD, R, TMG \rangle \}$$

This is E - P model. CD is event detector, process P is a triple set, and TMG is transaction management program.

How to create an agile e-commerce system? Using E - A model on active database by means of document [4], when merchant system can not meet the customer's requirement, it allows customer define events and monitor it in real time, once the defined event appears, condition detector sends message to trigger management program, it evaluates the condition and requests TMG creating a transaction and executes it, so that it can realize the agile service function of merchant system. Figure 1 is the e-commerce system with agile ability.

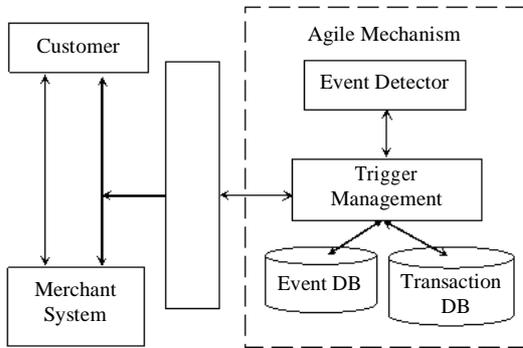


Figure 1 E-commerce system with agile ability

3. PROTOCOL DESIGN

3.1 Symbol

E : Entity. $E \in \{C, M, T, B\}$, C, M, T and B respectively stand for customer, merchant, TTP(Trusted Third Part) and payment gateway.

$A|B$: It is entity object. $A|B \in \{C, M, P\}$, C, M and P respectively stand for customer, merchant and payment gateway.

Req : It is goods information that is searched.

$A?B(X)$: It is that entity A ask entity B about information X .

$X+Y$: It is link of message X and message Y .

X, Y : It is series of message X and message Y .

$A"B(X)$: It is that entity A send message X to entity B .

$Prg.Opt[(Obj)]$: It is that procedure Prg executes operation Opt . Obj is operation object. $Prg \in \{CD, R, TMG\}$.

$D_{ASK}(X)$: It is signature to message X with secret key ASK of entity A .

$E_{BPK}(X)$: It is encryption to message X with public key BPK of entity B .

$H(X)$: It is abstract of message X , H is strongly one-sided Hash function.

OI : It is order information. $OI = (GName, GType, GNumber, dod)$, $GName$, $GType$, $GNumber$ and dod respectively stand for goods name, specification, amount and date of delivery.

GDB : It is merchant database.

$CertE$: It is certificate of entity E .

PI : It is payment information. $PI = \langle CAN, PW \rangle$

CAN : It is customer account.

PW : It is password of customer account.

MAN : It is merchant account.

$receipt$: It is merchant receipt.

CID : It is customer identity.

3.2 PROTOCOL DESCRIPTION

It is provided C, M, T and P have had verified certificate. C, M and B have registered from T , M has conserved the copy of certificate of T , T has conserved the copy of certificate of B . T acts as certificate, notarization, payment, arbitrate and time postmark in e-commerce protocol^[5], T plays the role of identity verifying, payment and evidence-protected in this protocol. This protocol consists of two sub-protocol, they are marked as Agility and Business. Generally, it executes Business protocol, Agility protocol is executed when merchant do not meet customer requirements.

Business :

$B1 : C^"M : (E_{MPK}(D_{CSK}(OI, Cert_t), H(OI)))$, C sends order information by way of signature to M .

$B2 : M$ decrypts the message with secret key MSK , and decrypts OI and $Cert_t$ with secret key CSK sent by C , and verifies truth of order information and identity of C .

$B3 : M^"C : (E_{CPK}(D_{MSK}(Cert_M, Cert_t, PI)$, M submits signature certificates of M and T and request payment information to C , certificate of T is preserved by M in advance.

$B4 : C$ decrypts the message with secret key CSK , and verifies certificates of M and T with secret key MSK sent by M .

$B5 : C^"M : (OP, CM, CT)$, $OP = H(OI) + H(PI)$, $CM = E_{MPK}(D_{CSK}(OI, H(PI)))$, $CT = E_{TPK}(D_{CSK}(PI, H(OI)))$, C sends payment information by way of dual signature to

M.

B6 : *M* decrypts *CM* with secret key *MSK*, and decrypts it again with secret key *CSK* sent by *C*, *OI* and $H(PI)$ are obtained, then $OI'=H(OI)$, $OP'=OI'+H(PI)$.

B7 : It creates $H(OP)$ and $H(OP')$ to verify their consistency, and compares them whether they are equal, if they are equal, it manifests *OI* is the order information related to *PI*, if not, it demands *C* send message again or cancel it.

B8 : $M^T:(OP, CT, KI)$, $KI=E_{TPK}(D_{MSK}(MAN, Cert_M, Cert_C, dod))$, *M* sends request payment information by way of dual signature to *T*.

B9 : *T* decrypts *CT* and *KI* with secret key *TSK*, and decrypts it again with secret key *MSK* sent by *M*, *PI*, $H(OI)$ and certificates of *M* and *C*, then $PI'=H(PI)$, $OP'=H(OI)+PI'$.

B10 : It creates $H(OP)$ and $H(OP')$ to verify their consistency, and compares them whether they are equal, if they are equal, it manifests *PI* is the payment information related to *OI*, if not, it demands *M* send message again or cancel payment.

B11 : Prepayment. *T* transfers *pm* to *MAN* from *CAN*, and sends successful message *prepaid* to *M*. $T^M: E_{MPK}(prepaid)$.

B12 : *M* decrypts it with secret key *MSK*, then sends payment time *dop* to *C*, $M^C: E_{CPK}(dop)$, $dop < dod$.

B13 : *C* decrypts it with secret key *CSK*, then $C^M: E_{MPK}(agreement)$, *C* agrees to the payment time.

B14 : *M* decrypts it with secret key *MSK* and gets message *agreement*, then merchant delivers goods to customer, and sends receipt to *T*, $M^T: E_{TPK}(D_{MSK}(receipt, H(Cert_M)))$.

B15 : Customer receives goods and verifies it is no problem, and sends message to *C*, $C^T: E_{TPK}(D_{CSK}(H(Cert_C)))$.

B16 :

(1) *T* decrypts them and gets message *receipt* and $H(Cert_C)$, and verifies identification of *M* and *C*, then sends payment information to payment gateway *B*, $T^B: E_{BPK}(D_{TSK}(PI, pm, MAN, Cert_T, H(Cert_T)))$, $Cert_B$ is conserved in *T* when merchant registers.

(2) *T* has only received message *receipt* from *M*, and verified identity of *M* is legitimate, or received unreal identity message from *C*, when payment time *dop* comes, *T* send payment request to *B*.

(3) *T* has only received message $H(Cert_C)$ from *C*, and verified identity of *C* is legitimate, *T* send payment request to *B*.

(4) *T* has not received any message within *dod*, it cancels payment automatically and informs *M* and *C*.

B17 : *B* decrypts message from *T* with its secret key *BSK*, and verifies signature identity of *T* with secret key sent by *T*, *B* formally transfers *pm* to *MAN* from *CAN*.

B18 : *B* returns successful message *success* of payment, $B^T: E_{TPK}(success)$.

B19 : *T* decrypts *success*, $T^C: E_{CPK}(finish)$, $T^M: E_{MPK}(finish)$, *T* sends finished message to *C* and *M*.

Agility:

A1 : $C^M:(Req)$. *C* inquires *M* for *Req*.

A2 : If *M* has *Req*, then $C^M:(E_{MPK}(D_{CSK}(CID, H(CID), OI)))$, $OI=(GName, GType, GNumber, dod)$, *GName*, *GType*, *GNumber* and *dod* respectively stand for goods name, specification, amount and delivery of date. *C* sends message with signature and encryption to *M*.

A3 : *M* decrypts message with its secret key *MSK*, then verifies identity of *C* with the secret key *CSK* sent by *C* according to $H'(CID)=H(CID)$.

A4 : Define E_u (Have *GName* in *GDB*), *M* defines event E_u .

A5 : *CD.Start*, Event detector *CD* starts.

A6 : *CD.Scan(GDB)*, *CD* scans *GDB*.

A7 : $CD^R:(E_{MPK}(E_u \text{ arises}))$, When defined event appears, *CD* sends the message to trigger management program *R*.

A8 : *Evaluate Cond(GType, GNumber, dod)*, Trigger management program *R* evaluates condition.

A9 : *R.Activate(TM)*, *R* activates *TM* when condition is given.

A10 : *TM.Create(t)*, *TM* creates relative transaction *t* and executes it.

A11 : $M^C:(E_{CPK}(OI))$, *M* sends message *OI* to *C*.

A12 : *C* decrypts message *OI* with its secret key *CSK* and obtains the order information. The protocol is end till now.

3.3 ATOMICITY AND SECURITY

Protocol should meet atomicity of money, goods and affirmable message^[6]. Known from B11 and B17, decrease of customer fund is increase of merchant fund, so it conforms to atomicity of money. Prepayment at step B11 meets both atomicity of money and preventing deceit deed of merchant and customer. Known from B15 and B16, customer begins to pay while he gets goods. Known from B14, B16 and B17, merchant deliveries goods while he receives payment money, so deed which customer gets goods without payment or merchant receives payment without delivery is not happened. This way of payment in getting goods conforms to atomicity of goods. Known from B14, B15 and B16, TTP asks payment gateway for payment and preserves transaction evidence when TTP has received receipt of merchant and identity message consent to pay by customer, this realizes atomicity of message who send it.

Truth to identity of transaction entity is important guarantee of protocol security. Certificate is the legal mark of identity, entities apply certificate authority for certificate. It needs to identify the truth of identity through delivering certificate in transaction entities. There is six times identification among four entities, *C* and *M*, *C* and *T* verify each other, and certificates of *T* and *B* are respectively conserved in *M* and *T* in advance, so *T* only needs to verify identity of *M*, *B* only needs to

verify identity of T . Process to identification consists of B1, B3, B8 and B16. Figure 2 shows the process.

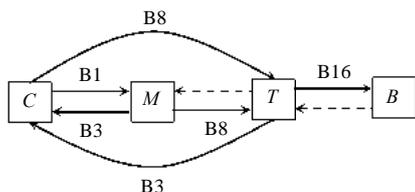


Figure 2 Identification process

Before formal payment, T needs to verify identity consistency of M and C . Certificates of M and C are sent to T at step B8, $H(Cert_M)$ and $H(Cert_C)$ are created, T compares them to these sent through B14 and B15, if they are equal, it shows identity of M and C are consistent and true.

Deceit means M gets C money by cheating without delivering goods or C gets goods without payment. To prevent deceit is one of security problems for e-commerce protocol. Obviously, deceit deed may happen in period of Business protocol. M may send a false receipt to T without delivering goods when protocol executes step B14, C does not send its certificate or send a false certificate to T when protocol executes B15. Considering success or failure of communication, there may have nine kinds of situations when protocol executes.

(1) M and C are both honest, they send message to T successfully.

(2) M and C are both honest, M sends message to T successfully, and C is failure.

(3) M is honest, C is deceitful, they send message to T successfully.

(4) M is honest, C is deceitful, M sends message to T successfully, C does not send message to T .

(5) M and C are both honest, M fails to send message to T , and C is successful.

(6) M and C are both honest, M and C fail to send message to T .

(7) M is honest, C is deceitful, M fails to send message to T , C does not send message to T .

(8) M is honest, C is deceitful, M and C fail to send message to T .

(9) M is deceitful and sends message to T successfully.

Step B16 can solve above-mentioned problems. For situation (1), step B16(1) solves it, step B16(2) solves situation (2), (3) and (4), step B16(3) solves situation (5), step B16(4) solves situation (6), (7) and (8).

TTP is an entity of prepayment, it preserves certificates of M and C and transaction information as evidence at the same time. Responsibility in protocol executing is traceable, it resists denying. For situation (9), M

deceives successfully, that is to say, M sends a false receipt to T , T pays only when signature certificate of M . If C does not get goods when *dod* expires, it can find out M is to blame.

4. CONCLUDING REMARKS

E-commerce protocol with agility can handle customer active request in real time, it makes merchant e-commerce system respond to a customer request agilely, and changes positive e-commerce into active e-commerce, and wins more opportunities for merchant.

Security to e-commerce includes three layers, they are physical layer, data layer and business layer. This protocol bases on business application layer. Signature guarantees security of business information. MD5 is a kind of common Hash to signature. Found by Wang Xiaoyun, MD5 exists "conflict", two files can create the same finger mark, this discovery makes law effect of current signature challenged^[7]. Universal used signature algorithm in e-commerce is worth to research.

Agent technology becomes more and more mature, it has wide prospect in e-commerce. Protocol introduced here is only tentative program, it does not introduce the content of intelligence search, negotiation, transaction agent and so on, and these valuable tasks are worth to study for us.

ACKNOWLEDGEMENT

This paper is supported by my tutor, Prof. Xiong Qianxing.

REFERENCES

- [1] Chunxiao L., Hui A., *E-commerce—from idea to action*, Beijing: Tsinghua University Press, 2001.
- [2] Aslam, Taimur. Dr. Dobb, "Protocols for E-Commerce", *Software Tools for the Professional Programmer*, Vol. 23, No.12, pp52-57, 1998.
- [3] Renee Gotcher, "Tools and Protocols for E-Commerce", *Information Security Technical Report*, Vol. 3, No. 2, pp34-40, 1998.
- [4] Yunsheng L., *Advanced Database Technology*, Beijing: National Defence Industry Press, 2001, 3.
- [5] Sihan Q., "TTP Roles in Electronic Commerce Protocols", *Journal of Software*, Vol. 14, No. 11, pp1936-1943, 2003.
- [6] Longxiang Z., "An Overview of Protocol Research in Electronic Commerce", *Journal of Software*, Vol. 12, No. 7, pp1015-1031, 2001.
- [7] Cryptograph of finger mark is found out, Signature is absolute security or not, http://www.sict.ac.cn/news/xwcx.asp?corp_code=0000001960, 2004.

B2B Electronic Markets: A Conceptual Model for Analyzing the Sources of Business Value

Chia Yao Lee¹, Brian Corbitt²

¹ School of Information Systems, Deakin University, VIC 3125, Australia.
chia.lee@deakin.edu.au

² School Graduate School of Management, Shinawatra University, Bangkok, Thailand, and
School of Information Systems, Deakin University, VIC 3125, Australia.
bc250604@yahoo.com.au

ABSTRACT

This paper describes research supporting the development of a conceptual model for understanding the sources of business value of Business-to-Business (B2B) Electronic Markets. Based on six case studies and an analysis of current literature, Aggregation, Matching, and Integration emerged as the three key sources of business value. The framework provides a structured and systematic approach for understanding various B2B Electronic Market models, and helps develop strategies to leverage these sources of business value.

Keywords: Electronic Markets, Business-to-Business Electronic Commerce, Business Value of IT

1. INTRODUCTION

In the wake of the Dotcom crash in 2000, interest in Business-to-Business (B2B) Electronic Markets have waned. Nevertheless, this should not be interpreted as a failure of the Electronic Market model in gaining traction amongst organizations and industries. On the contrary, Electronic Markets such as ChemConnect, Exostar and TheSeam.com have transacted millions, if not billions of dollars worth of transactions in the post-Dotcom days. In 2002, ChemConnect facilitated \$8.8 billion worth of transactions, penetrating up to 44% of the chemical industry players [1]. As Wise and Morrison argued, whilst the benefits of Electronic Markets are clear, the path to achieving them is anything but [2].

This paper describes the development of the conceptual model, using empirical data from six case studies, and an analysis of existing literature on Electronic Markets. The empirical data and literature analysis suggest *Aggregation*, *Matching*, and *Integration* are the three principal sources of business value associated with Electronic Markets. Aggregation, Matching and Integration are then lined-up against a series of business processes that typically underpins inter-organizational electronic commerce – (i) Information Gathering, (ii) Supplier Contact, (iii) Background Review, (iv) Negotiation, (v) Fulfilment, (vi) Consumption, maintenance and disposal, and (vii) Renewal [3].

2. CONCEPTUAL FRAMEWORK

Bakos defines Electronic Markets as Inter-Organizational Systems (IOS) that facilitate participating buyers and sellers to exchange information about prices and product offerings [4]. Many modern web-based Electronic Markets have evolved from intra-organizational database systems, simple electronic

distribution channels and electronic bulletin boards, such as TELCOT [5], ASAP [6, 7], and SABRE [8, 9], into complex Electronic Market systems that support a whole range of products and processes.

Current literature on Electronic Markets concentrates on the types of products and industries which have adopted Electronic Markets [10, 11] and the effects of adoption [9, 12, 13]. Kaplan and Sawhney proposed a framework for categorizing Electronic Markets according to the nature of transactions – spot vs. systematic purchases, and the types of products transacted – direct vs. indirect supplies [13]. Choudhury et al. and Lee evaluated in [10] and [12] respectively, the effects of Electronic Markets on product prices, and found no conclusive evidence to suggest that Electronic Markets lowered the prices of products. The effect of Electronic Markets on inventory levels was also examined in [12]. Again, minimal effects were observed.

Malone et al. regarded Electronic Markets as a structure that governs inter-organizational relationships according to the Transaction Cost Theory [14], asserting that the use of information technology will lead to a shift toward markets from hierarchies. Using a micro-economics approach, Bakos asserts that by reducing the search costs involved in inter-organizational trade, Electronic Markets increase market efficiency [4]. Taking a multi-disciplinary approach, Amit and Zott provided a holistic overview of various theoretical approaches towards the issue of business value of Electronic Business [15]. This study seeks to understand “how” and “why” organizations derive business value through an analysis of the sources of business value. While there are numerous sources of business value, the present study examines the three principal sources which have emerged from the case studies and analysis of literature.

Aggregation, Matching, and Integration have been identified as important in Electronic Markets [13, 16, 17]. The model developed in this study incorporates these three themes, using them to analyze and evaluate the activities supported by Electronic Markets. This study does not intend to test or prove the causality between Electronic Market adoption and the derivation of business value. Instead, it explores contemporary themes in Electronic Markets [2, 18].

3. METHODOLOGY

The objective of this study is to discover a better means of understanding how organizations benefit from engaging in B2B Electronic Markets and to enable this through a formal framework. Following the method prescribed by Yin [19] and Stake [20], data was collected from six case studies. A qualitative approach was adopted for this study to gain the rich data that emerges from in-depth studies [21]. Semi-structured interviews were conducted with 16 executives who were actively involved in the Electronic Market operations. Outcome from the interviews was matched with the findings from existing literature to form the conceptual model. The following is a summary of the six case studies conducted:

Case One

The organization involved is a finance-oriented company¹. It is a subsidiary of a multinational company located in the US. The organization uses online reverse auctions to source indirect supplies for its various business units. A sample of some of the indirect supplies it has sourced via reverse auctions [22] includes office supplies, accommodation, cleaning and courier services. The organization engages in reverse auctions with the objective of improving the supplier bidding process, increasing competition among the suppliers, and pooling together purchases for its different business units.

Case Two

This case involves an organization which set up an Electronic Market that catered for life science supplies. Members of the Electronic Market are able to buy and sell from that market. The sellers are mainly firms producing or retailing chemicals and apparatus, while the buyers are private laboratories and government agencies. At times, organizations that sell on the Electronic Market also buy from it. The organization that owned the Electronic Market had future plans to create new Electronic Markets in other industries.

Case Three

This case surrounds an Electronic Market for steel supplies. The steel Electronic Market caters especially for the Asia-Pacific region. The highly segmented steel

supplies industry in this region provided a great opportunity for the Electronic Market to link up with many smaller-sized steel suppliers that are geographically distributed. The Electronic Market for steel provided additional value-adding services, such as insurance, finance, freight and forwarding services to buyers and sellers. The concept was to provide a one-stop service for both buyers and sellers, which are made up of many small and medium businesses, which otherwise would not participate in Electronic Markets.

Case Four

The organization represents a branch of a leading computer component manufacturer. It uses an extranet system to coordinate with its distributors. The distributors are able to place their orders on the extranet and check on component availability. The extranet is extremely useful in streamlining the distribution chain in shortening the business cycle time, which is especially important in the computer component industry where obsolescence is costly and minimal inventories are kept. The extranet system allows the component manufacturer to gain access to real-time market information, and adjust its production capacity accordingly.

Case Five

This case involves an Electronic Market that was set up by a telecommunications organization for indirect supplies. The electronic market is targeted at its trading partners. It is aimed at creating an online community of organizations which could share each other's contact, increasing business between members of the community. Some of the goods and services which have been traded on the Electronic Market include office supplies, furniture, hotel accommodation and office equipment maintenance services.

Case Six

The organization involved represents a major healthcare product supplier. It maintains an online catalogue of healthcare products, such as hospital supplies, diagnosis machineries and equipment. The buyers are made up of hospitals, clinics, laboratories and specialized healthcare providers. This healthcare product supplier maintains close relationship with the major healthcare product manufacturers, and as such, operated as an intermediary between the manufacturers and the buyers. Buyers are able to place orders and check on delivery status electronically. Invoicing and electronic payments have been enabled with selected customers.

4. DEVELOPING THE CONCEPTUAL FRAMEWORK

Three steps were involved in the development of the conceptual model.

Step One involved identifying the three main sources of business value. The aim of this first step is to categorize

¹ Names of the companies are omitted at their request because of commercial considerations.

the different sources of business value which are relevant to B2B Electronic Markets. It is important to keep in mind that this study does not intend to test the strength and causality of Electronic Markets adoption and the derivation of business value. The purpose is to suggest the sources of business value based on themes which emerged from the six case studies. The cases were chosen from a variety of Electronic Markets from different industries to avoid the results from being limited to one form of marketplace only.

Step Two involved identifying the business processes facilitated by Electronic Markets. There are several ways for viewing the various business processes involved in a typical business transaction. For the purpose of this study, the Business Procurement Life Cycle model [3] was adopted. The seven stages of the business processes are: (i) Information gathering, (ii) Supplier contact, (iii) Background review, (iv) Negotiation, (v) Fulfilment, (vi) Consumption, maintenance, disposal, and (vii) Renewal. The objective of identifying these seven business process stages is so that the sources of business value can be studied depending on the business process it influences.

Step Three involved identifying the stakeholders. This final step involves identifying the stakeholders, so that the sources of business value can be specified and evaluated from different perspectives. The sources of business value may take on different meanings depending on the role of the stakeholder in the Electronic Market. Ideally, Electronic Markets create win-win situations for buyers and sellers. The parameters identified in Steps One, Two, and Three are then combined into a grid where the business processes are represented by rows, and the sources of business value, as viewed from different stakeholder perspectives, are represented by columns. It should be noted that there are numerous sources of business value that could be considered, and this study is by no means exhaustive in suggesting Aggregation, Matching, and Integration as the main ones. Rather, they appear to be the ones most relevant to B2B electronic markets.

5. DISCUSSION

The three significant themes which have emerged from the six case studies suggest their importance in the development of an analytic framework for B2B Electronic Markets. The indirect material procurement Electronic Market in Case Five demonstrated how the organization in question shared its business contacts with its trading partners, including allowing the trading partners to enjoy discounts and superior contractual terms that the organization has already negotiated with its suppliers. The pooling of this resource was used to encourage greater trading volume, creating a win-win situation for all the members of Electronic Market.

Stakeholders with similar interests pool their purchasing

power, resources and expertise via Electronic Markets. Aggregation is not new. It had occurred prior to the availability of Electronic Markets, for example, co-operatives were formed to help their members buy and sell, but with Electronic Markets, aggregation has further encouraged organizations to pool purchasing power, resources and expertise across geographical and temporal barriers. Aggregation is especially beneficial to smaller-sized organizations, which by themselves have too little influence and negotiating power. Aggregation is useful to the bigger organizations as it allows them to share their expertise with their trading partners, creating a closely-knit Electronic Market community.

The second theme which emerged from the six cases was that Electronic Markets enable buyers and sellers to seek out each other with greater ease. Electronic Markets often allow stakeholders to gain access to new market segments. The Electronic Market may also create a whole new market-space, playing the role of an intermediary. In the case of the steel Electronic Market in Case Three, it has opened up what was previously a highly segmented market. In a crowded market-space that has stakeholders with varying IT capabilities, the steel Electronic Market gave them access to the international steel market. The stakeholders were able to gain information about potential trading partners with minimal effort in advertising and tendering. Equipped with the information, buyers and sellers were able to find trading partners that could meet their requirements. This theme suggests that Electronic Markets may improve a market's efficiency, in that they link up sellers with buyers, reducing the effort of both parties in searching for each other.

Apart from enabling organizations to access new trading partners, it was also found that Electronic Markets allow existing trading partners to seek out new opportunities within existing trading relationships. As observed in Case Six, the healthcare product supplier was able to sell more types of products to existing customers. The online catalogue provided rich information on a huge range of products. The online catalogue is also capable of informing buyers of alternative products. If a required item is not available, then the online catalogue suggests the closest substitute. Were it not for such information on alternative products, the customer could have easily taken its business elsewhere.

In the case of the reverse auctions in Case One, it was observed that a bidder for a contract to supply services to the organization in question in a particular city was able to bid simultaneously for similar contracts to supply an organization in another city. Suppliers were able to seek out new business opportunities with existing trading partners via Electronic Markets. This is due to the fact that the reverse auction tendering system allows the organization to gain access to a greater number of bidders with marginal effort. Electronic

Markets enhance information exchange and sharing within a trusted business community, allowing buyers and sellers with specialized requirements and skills to seek out each other. As suggested by the channel manager of the extranet in Case Four, the Electronic Market system enables his organization to micro-market, gaining access to niche markets, to the extent of creating market segments which have only one customer. For sellers, Electronic Markets enable them to sell to buyers of highly customized goods. For buyers, Electronic Markets give them access to a bigger supply base capable of meeting their demand for customized goods. The organization in Case Four benefited from gaining up-to-date market information from its extranet. Customer orders determine the operations of the supplier, resulting in fewer inventories held. As such, the organization was more agile and flexible and could respond to changes in the market quickly.

The third theme which emerged from the cases relates to Electronic Markets enabling organizations to synchronize their business processes, improving coordination and collaboration with trading partners.

This idea of coordination and collaboration between organizations is not entirely new either. The Supply Chain Management concept describes how members of a supply chain share inventory and logistical data. In the Electronic Markets era, however, Integration relates to the synchronization of inter-organizational processes

that occur at the pre-transaction and the post-transaction stages. Electronic Markets can be used at the product development and design. Manufacturers can collaborate with suppliers via Electronic Markets to jointly develop products, while constantly receiving feedback from customers. The information shared among organizations may include those used for project management, product testing, and customer feedback and market research.

In Case Three, the steel Electronic Market provided facilities for insurance, finance, freight, and forwarding facilities to buyers and sellers. The Electronic Market is not limited to providing order-processing and invoicing facilities. Integration was demonstrated in Case One, where that organization used the Electronic Market to integrate its tendering and bidding processes. Although reverse auctions were mostly used for sourcing new supplies, the organization involved has even used reverse auctions to tender out contracts for the maintenance and disposal of computers and peripherals.

This illustrated how Electronic Markets integrate business processes at the post-transaction stages. Electronic Markets, it appears then, can function beyond supporting order-processing, invoicing, and logistics-related business processes. Table 1 below summarizes Aggregation, Matching, and Integration, as the emerging themes from the six case studies.

Table 1: Summary of case studies and emerging themes

Case Study	<i>AGGREGATION</i>	<i>MATCHING</i>	<i>INTEGRATION</i>
Finance Organization Reverse Auctions	Aggregating demand (pooling of purchases)	Electronic bidding for price discovery and determination	Integrating tendering & bidding processes
Life science Electronic Market	Aggregating demand and supply	Accessing new market segments, accessing new trading partners	Integrating tendering & bidding processes
Steel Electronic Market	Aggregating demand and supply	Accessing segmented market, accessing geographically distributed trading partners	Integrating freight, forwarding, insurance, payments
Electronic Distribution Channel	Little evidence	Sharing of time-critical market information	Integrating inventory management
Indirect Supplies Electronic Market	Aggregating demand and supply	Accessing new trading partners, leveraging existing trading partners for new sales opportunities	Little evidence
Healthcare Products Online Catalogue	Aggregating supply	Acting as an intermediary, allowing customers to access many trading partners	Integrating order-placing, invoicing, logistical processes

Data gleaned from the six case studies indicates that a particular source of business value influences different organizations at different business process stages. A possible explanation is that these Electronic Markets had different objectives, for example, the reverse auctions in Case One was a complementary procurement channel, whereas the life science Electronic Market in Case Two represented a new platform that brings together buyers and sellers within a market segment. At different stages of implementation

of Electronic Markets, the emphasis on the source of business value may vary also, e.g. new Electronic Markets concentrate on Aggregation, whereas established Electronic Markets concentrate on Integration.

6. CONCLUSION

The B2B Electronic Markets arena is still evolving. Economic rationalization brought about a survival of the

fittest scenario for Electronic Markets. Successful ones like ChemConnect and Exostar are gaining momentum while unsuccessful ones like Chemdex have faltered. A lot remains to be understood about the strategies underpinning successful Electronic Markets. While past research concentrated on the impact of information technology on business processes such as negotiation and fulfilment, this study demonstrated a framework that emphasizes how B2B Electronic Markets influence business processes such as information gathering, supplier contact, and consumption, maintenance and disposal. This is important as the benefits of Electronic Markets are not restricted to price and cost reductions. Inevitably those outcomes are a function of business process and any value generated in business process optimization.

The adoption of the proposed model for analyzing Electronic Markets-related problems or analysis of Electronic Market strategy by businesses will aid identification of value and formulate strategies based on real exemplars. Contributions of this study to an understanding of B2B Electronic Markets include that it: presents a structured way of understanding the benefits of Electronic Markets; helps identify future areas of opportunity; enables the analyst or business manager to view the business value from different stakeholder perspectives; enables researchers and practitioners to analyze and formulate strategies for Electronic Markets; provides propositions for future research; and enables the use of the model to analyze published cases of Electronic Markets.

REFERENCES

- [1] WSJ, "Top online chemical exchange is an unlikely success story", *Wall Street Journal*, URL: webreprints.djreprints.com/907660072246.html, 2004.
- [2] Wise, R., and Morrison, D., "Beyond the exchange: The future of B2B", *Harvard Business Review*, Vol. 78, No. 6, pp 86-96, 2000.
- [3] Archer, N., and Yuan, Y., "Managing Business-to-Business relationships throughout the e-commerce procurement life cycle", *Internet Research: Electronic Networking Applications and Policy*, Vol. 10, No. 5, pp385-395, 2000.
- [4] Bakos, J. Y., "A strategic analysis of electronic marketplaces", *MIS Quarterly*, Vol. 15, No. 3, pp295-310, 1991.
- [5] Lindsey, D., Cheney, P. H., Kasper, G. M., and Ives, B., "TELCOT: An application of information technology for competitive advantage in the cotton industry", *MIS Quarterly*, Vol. 14, No. 4, pp347-357, 1990.
- [6] Vitale, M., "American hospital supply corporation: The ASAP system", *HBS Case 9-186-005*, *Harvard Business School, Cambridge*, pp1-16.
- [7] Lee, C. Y., and Corbitt, B., "A stakeholder-benefit perspective of reverse auctions", *Proceedings of the 7th Americas Conference on Information Systems (AMCIS)*, Boston, 2001.
- [8] Hopper, M. D., "Rattling SABRE – New ways to compete on information", *Harvard Business Review*, Vol. 68, No. 3, pp118-125, 1990.
- [9] Christiaanse, E., and Venkatraman, N., "Beyond SABRE: An empirical test of expertise exploitation in electronic channels", *MIS Quarterly*, Vol. 26, No. 1, pp15-38, 2002.
- [10] Lee, H. G., "Do marketplaces lower the price of goods?", *Communications of the ACM*, Vol. 41, No. 1, pp73-80, 1998.
- [11] Markus, M. L., Banerjee, P., and Ma, L., "Electronic marketplaces in Hong Kong's trading industry", *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02)*, Institute of Electrical and Electronics Engineers (IEEE) Digital Library, 2002.
- [12] Choudhury, V., Hartzel, K. S., and Konsynski, B. R., "Uses and consequences of electronic markets: An empirical investigation in the aircraft parts industry", *MIS Quarterly*, Vol. 22, No. 4, pp471-507, 1998.
- [13] Kaplan, S., and Sawhney, M., "E-Hubs: the new B2B marketplaces", *Harvard Business Review*, Vol. 78, No. 3, pp97-103, 2000.
- [14] Malone, T. W., Yates, J., and Benjamin, R. I., "Electronic markets and electronic hierarchies", *Communications of the ACM*, Vol. 30, No. 6, pp484-497, 1987.
- [15] Amit, R., and Zott, C., "Value creation in e-business", *Strategic Management Journal*, Vol. 22, No. 6-7, pp493-520, 2001.
- [16] Bailey, J. P., *Intermediation and electronic markets: Aggregation and pricing in Internet commerce*, PhD thesis, Department of Electrical Engineering and Computer Science, University of Maryland, 1998.
- [17] Davenport, T. H., Cantrell, S., and Brooks, J. D. *The Dynamics of eCommerce Networks*, Cambridge, Mass., Accenture Institute for Strategic Change, URL: www.accenture.com/xdoc/en/ideas/institute/pdf/Dynamics_of_eCN_WP.PDF, 2001.
- [18] Kambil, A., and van Heck, E., *Making markets: How firms can design and profit from online auctions and exchanges*, Boston, Harvard Business School Press, 2002.
- [19] Yin, R. K., *Case study research : Design and methods*, Thousand Oaks, California Sage Publications Inc., 1994.
- [20] Stake, R. E., *The art of case study research*, Thousand Oaks, California, Sage Publications Inc., 1995.

Benefit Evaluation Model of E-Government Procurement

Mingzhu Jin, Xuping Jiang

School of Economic and Management, Tsinghua University, Beijing 100084, China
{Jinmzh, Jiangxp}@em.tsinghua.edu.cn

ABSTRACT

Since 1990s, Information Technology has developed at speed. With the development of IT and Internet, E-Government affair has become the focus of every government. As an important part of E-Government Affair, E-Government Procurement has a great influence on the development of economy. Based on the situation of Government Procurement of China, this paper states the importance of benefit evaluation of government procurement, and attempts to build up the benefit evaluation system model of Government Procurement in China, separately from economic benefit, political benefit and social benefit. Further more, some criteria for the information collection of E-government Procurement will be summarized.

Keywords: E-Government; E-Government Procurement; Government Procurement Agreement (GPA); WTO; Benefit Evaluation System; criteria system; Analytic Hierarchy Process (AHP)

1. INTRODUCTION

E-government is a kind of application information system realized by information technology, which is used to serve the detailed business, transaction, and conference of government. It is on the basis of the substantial government organization and resources in the Internet. Concretely speaking, E-government is a kind of behavior pattern of government depending on the Internet, which is closely related to the government management and society business service.

E-government is the result of Information Technology Revolution. In the face of the challenge of IT and Information Society, E-government development has become the focus of every country in the world. As result, the efficiency of government and service quality is improved by IT, at the same time the competitive power of the country is increased greatly to be against the challenge of Information Society through E-government. At present, E-government development are accelerated in all the countries in the world, which is not only important to improve the efficiency of government affairs and public service, but also especially important to improve the transparency and democratization of government.

Until to 2003, all office business has been put on the Internet in American Federation Institutes and Singapore government; government of Japan launched the project "new era" in February, 2000 and planed to perform all the government business during the period from 2003 to 2005; until 2002, 41% of citizens in U.K. were served mainly by the website of government institutes of U.K.

According to the report of world economic forum: the maturity degree of E-Government in China is 23%, which is still in the stage of building the basic frame. Three kinds of information system of E-Government

have been set up in our country: office system network among government departments; information sharing and real time online communication between all the functional government departments; information dual-communication between the government and citizens through the Internet. Some foreign academic institutions have put forward some evaluation standards for E-government, but in China there is still not a uniform, scientific evaluation criteria system for E-government, which draws attention of the academicians in China.

2. OVERVIEW OF GOVERNMENT PROCUREMENT

As one important part of the E-Government system, government procurement is the key node of public financial system, which is one of basic methods by which financial management and payment of the government is carried on under market economy system. According to government procurement agreement (GPA) in WTO, government procurement is defined as: the purchase and lease of the goods, service, projects of the central government and sub-government of member countries in WTO, and foundation of public facilities. Generally speaking, the procurement of central government takes on about 10-15% of GDP.

From 1990s on, with the development of IT and popularization of the Internet, government procurement on the Internet has become one of the main parts of government procurement revolution in many countries. The advantages of E-procurement are shown below:

- a. Time saving. During the whole E-procurement process, time is greatly reduced and procurement efficiency is greatly improved.
- b. Cost cutting. During the E-procurement, almost all the payment on the printing and posting can be saved, and workload is reduced, so that the cost of procurement is greatly

- reduced.
- c. "Open, justness, justice" principle sufficiently sticking to. The reduction of information asymmetry and corruption can make competition in government procurement market more sufficient and closer to "business in the sunlight".
- d. Effective and active government impression shaping. With the application of modern network in E-procurement, government shows aggressive and revolutionary spirit, which strengthens the belief of the citizens to government as result.

In China, government procurement is still a new thing. In the middle of 1990s, some provinces in China began to test the government procurement reformation, and it is Shanghai city that was first to try the reformation, and it is Shenzhen city that was first to set up the regional law for government procurement.

At present, government procurement is in the process of change from primary phase to entirely developing phase. In 2002, government procurement was supported by all levels of governments and parties, and received great achievement that the total procurement broke through 100 billion RMB, 100.96 billion. According to the statistic data from 37 provinces, the target of government procurement is 113.54 billion RMB in 2002, and the real payment is actually 100.96 billion, the cash saving is 12.58 billion RMB, ratio is 11%.

3. BENEFIT EVALUATION SYSTEM MODEL OF GOVERNMENT PROCURMENT

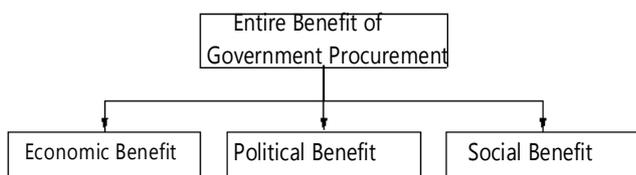


Figure 3.1: The structure of government procurement benefit

3.1 Economic Benefit Evaluation System

3.1.1 Evaluation Criteria

3/4 Capital Total of Government Procurement

This criterion reflects the scale of procurement of a country or a district. At the same time, it is the basis of international comparison, to measure the development degree and total level of government procurement. For example, in 1998, the capital saving ratio of government procurement is 13.38%, higher than 12.96% of 2000, but the scale of the later is ten times of the former, the total capital saving increased from 4150 million RMB in 1998 to 4.25 billion RMB in 2000. This comparison shows, total capital of government procurement is

highly related to economic benefit, which is the basis of economic benefit evaluation.

3/4 Ratio between procurement capital and GDP or Finance Expense

These two criteria reflect the importance of government procurement to national economy and finance expense. According to the experience of developed countries, the usual value of these ratios is between 10% and 30%, generally speaking, the more these ratios are closer to these numbers, the more benefit is obvious.

3/4 The Product and Value Framework of Government Procurement

Now there are three products for government procurement in the world: goods, project and service. Generally speaking, if all the products focus on goods, this fact shows that the government procurement of this country or district is still in the primary stage, although benefit is goods, but the scale is not large enough, the levels are not enough, as the result, there would not be obvious total benefit, just like China. If the distance among three kinds of products, or the later two are larger, the fact shows that government procurement goes into the higher stage, goes deep into all the parts of economic area. At this time, the scale of total government procurement is larger, and total benefit and framework benefit becomes more obvious.

3/4 Capital Saving Ratio

Usually this ratio for government procurement represents the difference between separate procurement expense and government procurement expense, divided with total procurement capital, but now people often use the difference between budget of government procurement and actual government procurement contracts to compute the capital saving ratio. The formula shows as following:

If P_{budget} denotes budget of government procurement, $P_{contract}$ denotes actual government procurement contracts, then

$$\text{Capital saving ratio} = \frac{P_{budget} - P_{contract}}{P_{budget}} \times 100\%$$

Generally speaking, the more this ratio is higher, the more the benefit of procurement is more obvious. Therefore, at present the high capital saving ratio in China illustrates the fact that the budget management in China has been too loose for long time, which also reflects the necessity of government procurement system. For long term, cost saving ratio of government procurement is the scientific basis for evaluating procurement benefit.

3/4 Development Speed

The formula is below:

If $P(t)$ denotes the procurement capital of t year, then

$$\text{Development Speed} = \frac{P(t) - P(t-1)}{P(t-1)} \times 100\%$$

This formula reflects the dynamic development of government procurement. On one hand, high speed illuminates great progress has been achieved on the procurement work; on the other hand, it illuminates former radix is too small, the procurement is still on the primary stage and has great room and potential to develop. For example, in China, there is 3.1 billion RMB of procurement in 1998; but in the following two years, the procurement developed in the speed of 300%, the total scale of government procurement expended greatly and same things happened to the economic benefit.

3.1.2 Economic Benefit Evaluation System of Government Procurement

Based on the statement above, Following is economic benefit evaluation system of government procurement.

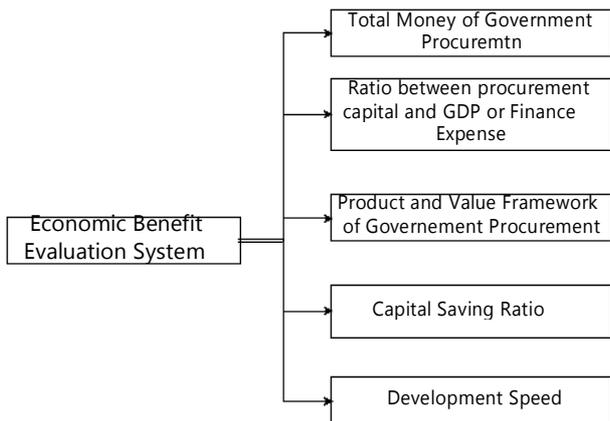


Figure 3.2 Economic Benefit Evaluation System of Government Procurement

3.2 Political Benefit Evaluation System

3.2.1 Evaluation Criteria

Political benefit of government procurement needs to be evaluated and analyzed by qualitative method. Government procurement carries on the policy mainly through accelerating social stabilization, protecting national industries to realize government benefit.

¾ Benefit Evaluation of accelerating social stabilization

Stabilization is the most important to politics, to maintain social stabilization is the greatest political benefit. Government procurement maintains the social stabilization through two ways: macroscopical adjustment and control; directly expand occupation.

Government procurement is an important tool to macroscopically adjust and control national economy. As the largest consumer, government owns the largest procurement scale in mass time. As a result, it would have great influence on macroeconomics operation. At the same time, government can adjust the procurement structure to adjust the industry structure, so such adjustment of economic period would benefit for stable increase of economy.

Government procurement can also increase the occupation. Occupation is a serious political problem, because it is directly related to social stabilization, related to people’s basic welfare. Especially during the period of economy stagnancy, government can expand social infrastructure to increase occupation.

Based on the analysis above, we can use two ratios to measure the benefit of accelerating social stabilization: economy development ratio and the ratio of employment by social infrastructure.

¾ Benefit Evaluation of protecting national industries

National industries involves with national benefit, which is also a political problem. Increasing government procurement makes China more import to the world. Usually we set up related laws for government procurement to protect domestic government procurement market.

United Nation has passed some related laws and requirements to clear the trade obstacle to guarantee trade liberty, and make sure that there is no discrimination for foreign suppliers. On the other hand, under certain condition, foreign suppliers are restricted to restrict take part in the government procurement in order to protect some key departments of the country from too much international competition. Foundation and improvement of government procurement system is absolutely necessary for protecting national political benefit.

Based on the statement, we attempt to use the number of related procurement laws and requirements, and the ration between the procurement from foreign suppliers and total government procurement that reflects the international degree of government procurement.

3.2.2 Political Benefit Evaluation System of Government Procurement

Based on the statement above, Following is economic benefit evaluation system of government procurement.

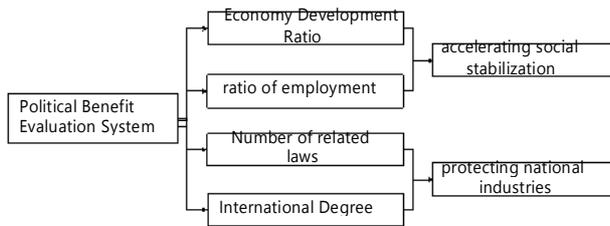


Figure 3.4 Political Benefit Evaluation System of Government Procurement

3.3 Social Benefit Evaluation System

3.3.1 Evaluation Criteria

Protecting environment, fighting against poverty, maintaining social order, belong to the benefit of public product. In order to realize these goals, government procurement needs to make some standards to evaluate its social benefit. According to welfare economic theories, increase in social benefit depends on the increase in social welfare, and increase in social welfare depends on whether personal welfare level is improved.

Although at present social benefit evaluation of government procurement is still neither complete nor correct, but from the analysis, evaluation result, we can conclude: although there are still many problems during the process of government procurement in China: just began to develop recently and the scale and scope is not large enough, the speed of development is high, the trend and progress of development is significant.

Actually, on maintaining the social order and clearing social ethos, China has achieved obvious social benefit through government procurement. Government procurement system makes enterprises realize that if one enterprise wants to win the government procurement market, it must depend on its credit and strength, it must train a sale force with high quality, which know everything about government procurement. As a result, many enterprises set up their own website to satisfy the requirements of E-government procurement, at same time make them famous and get more information of others, at the end improve the management level of their own.

Opening is the basis of justness. Justness can standardize the social order so that social benefit can be produced. The key to work out this problem is to complete and improve the present government procurement system further from the legislation.

At present, it is not obvious for government procurement to influence the social benefit from anti-poverty and environment protection, the main reason lies on the narrow scope of government procurement in China, there are still many products not included into the scope of procurement. At the same time, there are also some problems in the design of policy goals. Most government procurement focus on

economics goals, political goals and social goals still need more attention.

Based on the analysis above, we can illustrate the social benefit through two criteria: degree of justice, that is ratio between the number of the enterprises who own their own website among the suppliers and the number of the enterprises who are the suppliers of government procurement; degree of procurement public bidding, that is the ratio between procurement by public bidding and total procurement capital. These two formulas are following:

If N_w denotes the number of those enterprises who own their own websites among the suppliers; N_s denotes the number of the enterprise who are the suppliers of government procurement; $P_{bidding}$ denotes the procurement by public bidding; P_{total} denotes the total procurement capital.

$$\text{Degree of justice} = \frac{N_w}{N_s} \times 100\% ;$$

$$\text{Degree of publicity} = \frac{P_{bidding}}{P_{total}} \times 100\% .$$

3.3.2 Social Benefit Evaluation System of Government Procurement

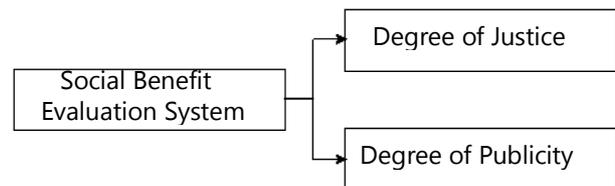


Figure 3.4 Social Benefit Evaluation System of Government Procurement

4. MULTICRITERIA EVALUATION METHODS INTRODUCTION

The multicriteria evaluation is widely used in the social life, recent years there are many multicriteria methods emerging, mostly used are AHP (Analytic Hierarchy Process), DEA(Data Envelope Analysis), PCA (Principal Components Analysis), and Expert Analysis. From benefit evaluation system, we can see that there are many criteria to evaluate benefit of government procurement, and AHP is the most suitable method for benefit evaluation of government procurement.

In the 1970s, AHP was invented by T. L. Satty, who was an American expert of operational research. AHP has been used to system analysis and strategy research in the area of economy and politics, with the characteristic of practicability, terseness and validity. The process of AHP includes problem conceptization, hierarchy structure foundation, consistency judgment matrix formation and so on, consistency judgment matrix formation is one of the key steps. This method can be used to research more complex multicriteria decision

problems, quantify the qualitative factors, to some extent test and reduce the subjective influence, and make the evaluation more scientific. In 1982, Mr. Xu shubai introduced it into China.

When the entire benefit of government procurement is evaluated by using AHP, first step is to separately, evaluate economic benefit, political benefit and social benefit, then get the values of each benefit; second step is to give weight to each benefit according the opinion of the decision maker, at last add up these weighted values, then get the final evaluation value. Decision makers can compare evaluation values of some years to give some advices for developing and improving government procurement work.

5. INFORMATION COLLECTING OF GOVERNMENT PROCUREMENT

From the analysis above, some information needs to be collected to convenience the benefit evaluation of government procurement. Such information can be divided into two types: statistic information, statement information. For statistic information, there is horizontal information, longitudinal information, annual statistic and so on, in the form of digital information. On the other hand, statement information is mainly the information that describes the characteristic of government procurement behavior, in the form of letters. Based on the former analysis, we get the a brief summarization of information collecting of government procurement:

Table 5-1 Information Collecting

Statistic information	Statement information
Monthly procurement	Type of procurement product
Annual procurement	Import or export
Finance expense	Degree of justness
Budget of procurement	Degree of publicity
Actual contracts sum	----
Laws of procurement	
District procurement	

During the process of government procurement, the information mentioned above should be paid more attention in order to evaluate the benefit of government procurement efficiently, effectively, conveniently.

6. CONCLUSION AND FUTURE RESEARCH WORK

As an important part of E-Government, E-Government Procurement has a great influence on the development of economy. Government procurement is one key node in the public finance system, which is one of the basic methods to perform the national finance management and payment in the system of market economy. Because

the characteristic of Government procurement is "justness, justice, open", it can greatly prevent the malversation, such as some personal operation and corruption, as a result usually the government procurement can be called as "project in the sun". Normative government procurement has the important meaning to the incorruption in the government. How to realize E-government, and how to use present information technology to evaluate the whole benefit of government procurement, that is the present focus of government affairs.

As the situation of global government procurement shows, E-government procurement is inevitable, which is the direction of government procurement reformation. In term of the general rules of E-government procurement, together with the fact in China, the basic design of E-government procurement is: strengthen the research and programming of E-government procurement, establish the mutual mechanism of government procurement reformation with E-government procurement step by step.

In China, the benefit evaluation of government procurement is still on the primary stage and still in the process for groping. In fact, most present evaluations are focus on the economic criteria, not social criteria and political criteria, and ordinarily strained to a few special criteria for quantity analysis. On the other hand, the result of evaluation is often affected by the value judgment, knowledge, experience of person who is making the evaluation, as well as by the quantity of the material and the relationship between the person making evaluation and the object evaluated, what is mentioned above will finally influence the justness and objectivity of the evaluation results.

This paper attempted to set up the evaluation system of the whole benefit for government procurement, separately from economic benefit, social benefit and political benefit three parts to explain the source of government procurement benefit. According to the different type of benefit criteria, different quantity and qualitative methods are used to evaluate the benefit: concretely speaking, for economic benefit, measure criteria are used; for social benefit and political benefit qualitative criteria are used.

The future research work will focus on exploration of the evaluation method for the enterprise procurement in or out of China, and those of government procurement in developed countries, some reasonable advice from the above will be provided for building a China-fitted evaluation criteria system according to China's situation. Scientific, reasonable, standardized benefit evaluation system for government procurement can make the evaluation process more easily, operation more conveniently, the time for evaluation less, at the same time, the lessons and experience of procurement can be summarized in time, as a result government

procurement becomes more transparent and more "sunlight".

REFERENCES

- [1] Xiaoqun He, "Modern *Statistic Analysis Method and Application*", People University of China publishing company, P281-315,1998
- [2] Daoyou Chen, "Benefit Finance: government procurement", economic science publishing company, Vol. 1,2002
- [3] Mingzhu Jin,Xuping Jiang, Influence of E-Government Procurement to Economy in China, "Chinese MIS",Vol.3, 2004
- [4] Xuping Jiang, "E-commercial and network marketing", Tsinghua University publishing company,Vol. 7, P169-180,1998
- [5] Yunquan Hu, Yaohuang Guo, "Operational Research", Tsinghua University publishing company, P387-398,1998
- [6] Ping Li, "Improvement of E-government and Public Service", *Management Science*, Vol.3, 2004
- [7] Xiaochuan Liu, Qinghua Wang, "Government Procurement Under economic globalizatioin", economic management publishing company, 2001
- [8] Hua Jiang, Guoan Gao, Dongliang Liu, "Design and Realization of Multicriteria, Multiobject Decision Evaluation System", *System Engineering Transaction*, Vol.4, 1997

E-Business Value Process Modelling

Baolin Wu, Li Li, Yun Yang

CICEC - Centre for Internet Computing and E-Commerce
Faculty of ICT, Swinburne University of Technology
PO Box 218, Hawthorn, Melbourne, Australia 3122
{bwu,lli,yyang@it.swin.edu.au}

ABSTRACT

E-business development is a considerable complicated task because the underlying logics of e-business and its new processes that are originated from implementing enhanced information technologies to streamline business performance introduce many complex issues. One of the difficulties is to capture the dynamic aspects of e-business that can be used for monitoring the business performance, in a way that could be helpful for the business adaptation to meet competitive advantages. Among many dynamics of e-business, the value system is the most interested one that has recently been addressed. However a value system is at the strategic level with no formal approaches for its representation, which introduces a gap between system modelling and implementation in the e-business development. In this paper, we will investigate a so called value process that can be not only used for value system modeling, but executed for simulation of the resulted model. For the purpose of value process modelling, we will adopt the process algebra approach, which will be integrated with others such as workflow in our modelling environment.

Keywords: E-business, value system, value processes, e-business modelling.

1. INTRODUCTION

There are considerable difficulties in modelling e-business and its processes, because e-business dramatically changes traditional business value systems and the patterns of action and interaction that bring together processes from different companies, sites, suppliers, distributors, service providers etc, as well as the new processes emerging from business needs to support the collaboration, coordination and cooperation. The e-business era has introduced many new paradigms, that has given a rise to a need for appropriate tools and modelling mechanisms that support for the better understanding and the decision making of the business and its process development.

An e-business is considered here as the integration of business rules, a viable trading mechanism and associated processes into a business approach that leverages the network (Internet) as its medium of transaction, it is thus composed of the characteristics of the definition of the offered products and services, the targeted customers, identification of the involved business partners and their roles, and the value propositions among business actors. With this context of e-business, the business strategy analysis will be, not only in business itself but in business process as well, an important element in e-business development. The business strategy analysis is a methodology of configuring the business value system that consists the value architecture and the value dynamic at run time. For the given business strategic plans and goals, the processes of developing a sound value system and value process underlying the value system become vital for e-business to be successful.

Recently the value based approaches have been advocated in e-business modelling [1, 4, 6, 8, 17, 23]. In these modelling proposals, e-business modelling is about the value and value operation for what is offered by who to whom in term of value, while e-business process modelling is the concerns of how the value offerings are selected, negotiated, contracted and fulfilled operationally [3, 4, 7, 21]. In [5, 6], an e^3 value based approach has been proposed that rigorously describes value exchange among business actors and formalises a value exchanging architecture based on requirements engineering and case mapping techniques. Because of the "soft" and complex nature of e-business, the description of e-business modelling is remained at the conceptual level, some graphical tools have been developed for visualising the business actors, value flows and value relationship among the actors, which can help in e-business value system development [4, 5]. However, the lack of computation mechanisms in the conceptual model of the value system causes the difficulty for simulation and feasibility testing. In this paper, motivated by e-business modelling demands, we propose a novel value process modelling approach that brings the e-business modelling and its process modelling together in a single form. With the computable for simulation in our mind, we adopt the process algebra [2, 11] as our modelling language that addresses two fold in e-business modelling - static and dynamic characteristics of the e-business value system. The value process of e-business is nothing but the values and operators that model what and how the value is created and transmitted within e-business with respect to the strategic plans and goals.

This paper is to study the value process modelling, and examines how to facilitate the modelling results to

derive e-business processes. The remainder of this paper is organised as follows. Section 2 presents the background of e-business modelling by reviewing the requirements in e-business modelling demands, and addressing the important relationship between the value system and the e-business process. In section 3, the value process is discussed using the adopted process algebra. In section 4, we give an example to show how the value process can be used in e-business modelling and development. Finally in section 5, the conclusion is given, and our further considerations for the remaining issues are addressed.

2. BACKGROUND

In this section, we briefly review the related literature in e-business modelling from the value point of view, and examine the value system of e-business, which gives the background for our development of the value process.

2.1 Value Based e-Business Modelling

An economic activity is directly toward the creation of value. A business is considered superior comparing to others, if that business offers a superior value to its customers in the competitive marketplace, and gains benefits from it. A value is an economic concept, not primarily measured in physical performance attributes, rather what a buyer will pay for value object - product, service, information and knowledge. A description of these related strategic value plans is *Value Proposition*. The constituents of a value proposition may include the clear answers of questions such as

- What are the contents of value to be offered?
- Who are the end users of the value?
- Where are the end users?
- What are the benefits for the end users, partners and itself?
- Why ought the end users accept the value offered?
- What the benefits the business can gain by offering the value?

Value proposition can be understood as the statement of benefits that are delivered by the business to its external constituencies, and that are gained to the business itself. However, the statement of a value proposition is often an informal one that can only be understood by the experienced business modellers. It is difficult to be understood by those who carry the development of the e-business processes and softwares. It is preferable that a formal approach defines a value proposition. Inspired by the works of several researchers in business modelling [4, 10, 12, 21], the value proposition may be described by four attributes including customer needs, value creation, value level and value price level. These attributes are closely related to the value articulation. The customer needs include the concerns of usefulness, risks and effectiveness of the value offered to the

customers. The value creation is the way of how the value can be created at the moment of the value life cycle of value object creation, consumption, renewal and transfer. The value level is the measurement of the value comparing to its competitors in scales of similarity, innovative imitation, excellence and innovation. And the value price level is the proposed price by comparing to one or all of the competition in the scale of free, economy, market and high-end. We note that defining a value proposition is the first step in e-business development, which requires business strategies that clearly point out the attributes of the value proposition. A value proposition, however, only provides a broad view about what the values of the business are to customers and itself. We can not see how the business operates the values at this level of modelling phase, which a value architecture is required that describes the partners' value relationship, capability, competence and resources for fulfilling the value propositions.

The value architecture addresses the problems such like who are involved and what the roles the actors play in the business, what relationships are among actors, and what activities and resources are needed to fulfill the promised benefits to the customer in an efficient way. The value architecture has been explored by several researchers, for example the e^3 provided a model of the value architecture that focuses on value and value exchange [5]. Generally speaking, the value architecture can be regarded as a value system that is a connected series of organisations, resources, and knowledge streams involved in the creation and delivery of value to the end customer. It is centralised about the value creation process and the problems of how to anticipate customers' and partners' processes by using Internet technology. Value systems integrate the business activities, from determination of customer needs through product/service development, production/operations and distribution. For such integration, it is required that the processes underlying the value system have to be explored. We consider that the value system is the aggregation of value processes that need to be modeled for e-business development.

2.2 Value System and E-Business Processes

To approach each business opportunity, adaptive business processes are required for continuously restructuring value systems. The relationship between the value systems and e-business processes is that the former provides the guidelines for how the e-business process should be in order to offer better values to its customers, while the business processes support the value systems such that the valid processes are facilitated to fulfill specified requirements of the value systems. Figure 1 shows the relation between them in the e-business development life cycle. However in addition to the conceptual level of the value systems in which might be difficult to be executed for simulation, the current approaches separate these issues that

introduce the gap that gives a rise to the transitional difficulty from a value system to a process system and vice versa. We need a modelling mechanism that supports the understanding of e-business and system development. We found that the difficulty is laid by the representation of the value systems, in which the dynamic aspects of the value systems can not be captured. Since a business process is an operational system that directs the business activities in steps, the conceptual model of the value systems might not be able to reveal the insights of the dynamics in business processes. A different notion is required for value system that bridges the gap of the value system and the e-business process. The value process, discussed in the following, is proposed to this challenge for e-business modelling, and it provides the syntax and semantics of a value system to capture the dynamic nature of the e-business systems.

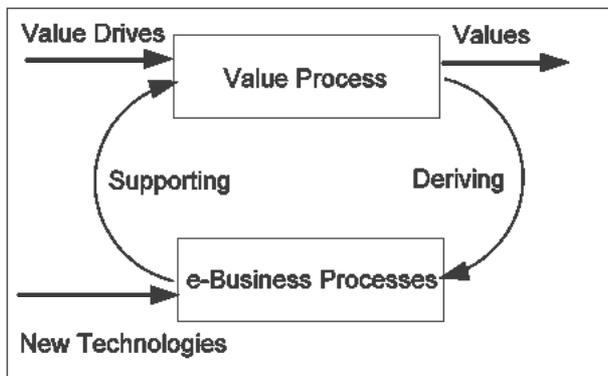


Figure 1: The dynamic relation between value process and business process.

3. PROCESS MODEL OF VALUE SYSTEM

Most of e-business modelling methods are text or graphic based that can explain the value system but provides the drawbacks in depicting the dynamics. Our modelling approach takes from the process point of view that provides the semantics of the value system in the sense of the formalisation of the value system that the modelling result can be executed for simulation. In dealing with the value process, our modelling notion adopts the process algebra for business applications such like concurrency and synchronous/asynchronous with respect to value. At the high level of value process, the primitives are agents and the interaction among them. On another hand, at the low level of value process (business process), the primitives are processes with the actions and activities in value creation.

3.1 Value Process Description

Two types of actions are ‘input’ and ‘output’ for a given ‘value channel’. Let $x, y, z \in V$ be a set of value objects, and $a, b, g \in C$ be a set of value channels. The prefix actions for a process are as

$a(x)$: The action of inputting the value x along the value channel a .

$\bar{a}(x)$: The action of outputting the value x along the value channel a .

We denote the prefix action set, A_v , which means that each action in the set is either input or output with the specified values and channels. Value process primitive P is an activity that acts on the value via the channel, and has its continuation of the process, denoted as $v.P$. In addition to the value process primitive, a set of operators expresses the composing rule of the primitives that form a complex process. Let $P, Q \in \mathbf{P}$ a process set, symbolically, the syntax rule of a value process is defined as

$$P ::= v.P \mid Nil \mid P + Q \mid P \mid Q \mid (rx)P .$$

where $v \in A_v$, a prefix action of input or output. The semantics of the value process syntax are intuitively understood as follows:

- Process $v.P$ defines a process with the value action prefix and continuation.
- Process Nil defines a value process that is no longer active to its environment.
- Summation process $P + Q$ represents an alternative of two independent processes that can be selected. Either P or Q is performed by the process at a time, but not both. For instance, getting the information of a value object either by searching on the Internet or reading from a newspaper.
- Composition process $P \mid Q$ represents two independent processes that are value dependent or independent parallel processes. When dependent, the process is interacting via its channel a and value x such as $a(x).P \mid \bar{a}(x)Q$, meaning that P and Q are dependent processes with respect to value prefix a , P acts offering value x , and process Q performs a complementary action receiving a value for the value process.
- Restrict process $(rx)P$ means the value process restricts its process on value x .

The notation of the value process above gives the process with the value and the channel as activity parameter. It is easily extended into multi-parameters expressed as $(v_1, v_2 \dots).P$, where $v_1, v_2 \dots \in A_v$, the syntax rule for a value process remains unchanged. The description of dynamic aspects of the value process is achieved by defining the deduction rules that govern the process at run time. Just as in the traditional process algebra, there is only one deduction operator \rightarrow over process; $P \rightarrow P'$ means that P can be transformed or evolved into P' by a single computational step. In

general, the interaction between two value processes is defined according to the following inference rules:

Interaction : $\bar{a}(x).P | a(y).Q \rightarrow P | Q\{x/y\}$.

Parallel : $P \rightarrow P' \Rightarrow P | Q \rightarrow P' | Q$.

Restriction : $P \rightarrow P' \Rightarrow (rx)P \rightarrow (rx)P'$.

Structure : $Q = P, P \rightarrow P', P = Q' \Rightarrow Q \rightarrow Q'$

The process algebra has its richness of mathematics. However we are interested in the use of the notion for value system modelling. By adopting it for the process of the value, we arrive at the value process, which we will intend to use for value system modelling.

3.1 Value Process Modelling

Based on the value process, we discuss how the notion can be used for value system and business process modeling. In e-business modeling, the value proposition is the first step getting into understanding the logics of e-business. We assume that the value proposition is articulated for use, and the information about values, actors involved, the competences and resources required of e-business are available. To represent a value system that connects these resources together, the value process is used to describe the actors and roles, and the value relationship among the actors.

A value process agent, \mathbf{A} , is a value process that offers or receives the value and continues the process. We note that an agent may be a primitive value process, $\mathbf{A}(v) = v.P$ or a composed one. In general, the agent may take many parameters as its prefix actions as required, $\mathbf{A}(v_1, v_2 \dots)$, while the value relationships are described by how the value is offered and received among these agents, and what value process are required for each agent to fulfill the value relationship.

We use the value process agent to represent the business actor. The parameters that agent takes are considered as tasks that agent will perform. An agent is an independent entity. The connection of agents is done by matching the dependency of the tasks dedicated to the agents. For example, assuming that there was three agents denoted as $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3 \dots$, each agent performs the tasks such as $\mathbf{A}_1(v_1, v_2, v_3), \mathbf{A}_2(v_2, v_3), \mathbf{A}_3(v_1, v_2)$, where v_1, v_2, v_3 is the task set. Then

$$\text{System} = \mathbf{A}_1(v_1, v_2, v_3) | \mathbf{A}_2(v_1, v_2) | \mathbf{A}_3(v_2, v_3)$$

It is an abstraction of the value process system, which only states the agents and their roles in terms of tasks. Elaborating this abstraction, we can examine how agents interact by expending the value process into the details with respect to tasks that an agent performs.

For example, we can specify the channels between the agents in term of the tasks.

$$\mathbf{A}_1 | \mathbf{A}_2 | \mathbf{A}_3 \rightarrow \bar{a}(x)P_1 | a(y)P_2 \rightarrow P_1 | P_2\{y/x\}$$

Here we assume that $\bar{a}(x)P_1 \in \mathbf{P}_1$ and $a(y)P_2 \in \mathbf{P}_2$ are two process sets that agents \mathbf{A}_1 and \mathbf{A}_2 will carry out to fulfill their tasks. The transformation of steps indicates the processes required for the value transmission between agents. Actually, we see that the detailed value processes can be regarded as the business processes, which show how the value is processed, created and added in each process step. We do not go further for the business process here as the purpose of the value process is to model the value system. The process notion discussed in the previous section provides enough information that supports the business process development, which will be our further interests in e-business modelling.

The financial structure of the value systems is one of the important modelling issues. For a given value system, we can inspect the financial aspects of the value process by estimating the cost and revenue sources, aggregating for the overall of the agent financial performance. We do these within the value process by defining a set of value calculators that estimate the cost of each process at the modelling phase, and compute the running cost and the revenue at the simulation phase in order to estimate the profit for each agent.

$$\text{Benefit} = \text{Revenue} - \text{Cost}$$

We argue that the financial aspect of e-business model can not be measured statically. The revenue sources and costs of the value system are dependent on running the system in which the transaction is made. The value process has detailed each value transaction between agents, thus it is possible to estimating the financial aspect of the value system at the simulation phase.

To this end, we believe that business knowledge and skills in strategic analysis are the critical and the most important component for the issues such as value capturing, activity identification, partner selection and pricing in designing an e-business system [10,13]. These issues are separated from the value process since the focus of value process modelling is to exploit the ways of how the value is processed in terms of creating, adding and distributing value and revenue among business actors. Developing a value system for business innovation or a new business idea is to capture the values to the customer and customer's values, which is a very sophisticated business process. We do not concern the value capturing; rather assume that there is the related information about the values that have been identified and available for the value process. To summarise the value process modelling, the following steps are considered:

- Articulating the values for a new or innovated business idea.
- Constructing value proposition of e-business.
- Investigating the value relationships between stakeholders, partners and customers.
- Constructing the value process system for a given value proposition.
- Estimating the financial aspects of the value system.
- Revisiting the value system for value refinement if necessary;
- Developing the business process to support the value process system.

5. AN EXAMPLE

We look at an e-business “brokage model”[15], and examine the value process in the model. Then we inspect how the value is processed in the model. The brokage model has ten sub-models including Marketplace exchange, Buy/Sell fulfillment, Demand collection systems, Auction broker, Transaction broker, Bounty broker, Distributor, Search agent, and Virtual Mall. Three process agents involved in the model, namely Seller, Buyer and Broker. The “flow-graphic” of the auction broker sub-model is shown in Figure 2.

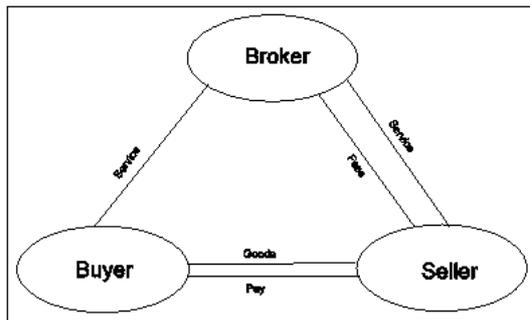


Figure 2: An auction model of e-business

In this auction model, the “Broker” provides the Internet services that a “Buyer” or “Seller” can use for the information of the goods and buying and selling the goods. On the high level of the business description, the Broker does the activities of auction by taking the information of goods from the Seller who wants to sell and auction the goods on behalf of the Seller, it charges the Seller a listing fee and commission scaled with the value of the transaction. On other hand, the Broker interacts with the Buyer according to offering and bidding rules, and finally fulfills the goods including price, delivery, etc. Three value process agents, involved in this model, are Broker, Seller and Buyer. A value set that describes the value relationship among the process agents are goods, money, service and information. Here we do not model the channels that are assumed as the Internet, and we only show the value system of auction and the business processes that the agents must perform.

The auction model is defined as

$$\text{System}_{\text{Auction}} \stackrel{\text{def}}{=} \text{Buyer} | \text{Borker} | \text{Seller}$$

The tasks that each agent will perform are

$$\begin{aligned} \text{Broker} &= \text{Brocker}(s1, s2, lf, af), \\ \text{Buyer} &= \text{Buyer}(s1, g, gp), \\ \text{Seller} &= \text{Seller}(s2, lf, g, gp), \end{aligned}$$

where $s1$ and $s2$ present Broker’s offered services to Buyer and Seller respectively, and lf, af, g, gp means the listing fee, auction fee, goods, goods payment respectively.

In the auction model defined above, the system shows that the agents and the value relationship between them. Next we define the processes for each process agent:

$$\text{Borker}(s_1, s_2, lf, af) = \overline{s_1}.B_1 + \overline{s_2}.B_2 + \underline{lf}.B_3 + \underline{af}.B_4$$

$$\text{Buyer}(s_1, g, gp) = \underline{s_1}.C_1 + \underline{g}.C_2 + \overline{gp}.C_3$$

$$\text{Seller}(s_2, lf, g, gp) = \underline{s_2}.D_1 + \overline{lf}.D_2 + \overline{g}.D_3 + \underline{gp}.D_4$$

where $B_i, i=1,2,3,4$, is Broker’s process, $C_j, j=1,2,3$ and $D_k, k=1,2,3,4$ are the processes of the Buyer and Seller with the prefix actions specified. If an interaction occurs, for instance between Broker and Buyer, at a service step, then the system is evolved into

$$\text{Buyer} | \text{Borker} | \text{Seller} \rightarrow \overline{s_1}(x).B_i | \underline{s_1}(y).C_j,$$

which indicates that the Broker and Buyer are interacting via the service channel. For the negotiation processes between the Buyer and Seller and negotiation service provided by the Broke, we can go further to detail the processes implementation. However, ot os unnecessary to give the trivial list of the processes.

The example of e-business modelled here by the value process shows three aspects. The value process gives the elements for the representation of the concepts of the value system, it glues the value system and business process together and the changes in the value system can be easily modelled to adjust the underlying process. The model can be simulated for reducing the risk of business development. For a particular e-business case, the value process should be made in a unique way for the competitive advantage. A value process aims at what and how the value is created for the targeted customers. The main benefit of a value process, from the competitive advantage point of view, is that it can be simulated, tested and reconfigured to meet the business strategic goals. For an identified value, how the value is processed among the business actors is determined by value-added analysis - an engineering pattern that focuses on eliminating activities that do not add value,

which has not modelled, but will be considered as our further investigation.

6. CONCLUSION AND FURTHER WORK

In this paper, we have proposed a novel e-business modeling approach. The core of the approach is a value process. The modelling approach is different to others: First it is a value driven approach concerning the values to the customer, then a value process focuses on what activities are required and how the values are created, finally the value system is configured through agents sharing the activities in the value process. We have pointed out that the business strategy is one of the most important components in constructing an e-business system. The business strategy analysis must be carried out on each phase of business modeling to build a competitive business system. An example of Brokerage model is given to demonstrate the modeling approach. Our further work will be to develop a modeling environment to support the modeling methods.

ACKNOWLEDGEMENT

Work reported in this paper is partly supported by Swinburne Vice Chancellor's Strategic Research Initiative Grant 2002-2004 for project "Internet-based e-business ventures".

REFERENCES

- [1] Amit, R., C. Zoot, "Value Creation in E-Business", *Strategic Management Journal*, 22: 493-520 (2001).
- [2] Baeten, J., editor. "Application of Process Algebra", v.17, Cambridge University Press, 1990.
- [3] Eriksson, H., P. Magnus, "Business Modelling with UML - Business Patterns at work", John Wiley & Sons Inc, 2000.
- [4] Giaglis, G., J. Paul, G. Doukidis, "Dynamic Modelling to Assess the Business Value of Electronic Commerce", *International Journal of Electronic Commerce*, 3 (3), pp. 35-51, 1999.
- [5] Gordijn, J., M. Akkermans, J. van Vliet, "Value based requirements creation for electronic commerce application", *Proceedings of the 33rd Hawaii International Conference On Systems Sciences*, CD-ROM, IEEE, Januray, 2000.
- [6] Gordijn, J. J.M. Akkermans, "Designing and evaluation e-business models", *IEEE Intelligent Systems - Intelligent e-Business*, pp.168-178, vol.16, no.4, 2001.
- [7] Gulati, R., N. Nohria, A. Zaheer "Strategic Networks", *Strategic Management Journal*, 21: 203-215, 2000.
- [8] Keeney, R., *Value Focused Thinking*, Harvard University Press, 1996.
- [9] Lamarque, E., "Linking Value Chain and Competence: Evidence from the French Banking Industry", *The 5th International Conference on competence-Based Management*, Helsinki, 2000.
- [10] Magretta, J., "Why business models matter", *Harvard Business Review*, Vol. 80, No. 5, pp.3-8, 2002.
- [11] Milner, R., *Communication and Concurrency*, Prentic Hall, 1989.
- [12] Osterwalder, A., L. Sarra, P. Yves, "An Ontology for Developing e-business Model", *DSIage 2002*
- [13] Porter, M., V. Millar, "How information gives you competitive advantage", *Harward Business Review*, 63(4), pp.149-160, 1985.
- [14] Porter, M, "Strategy and the Internet", *Harward Business Review*, 79(3), pp.62-78, 2001.
- [15] Rappa, M.m "Managing the digital enterprise - Business Models on the Web", <http://ecommerce.ncsu.edu/business/models.html>, 2004.
- [16] Slywotzky, A., *Value Migration: How to think several moves Ahead of the competition*, Harvard Business Press.
- [17] Stabell, C., O. Fjeldstad, "Configuring value for competitive advantage: on chains, shops, and networks", *Strategic Management Journal*, 19, pp.413-437, 1998.
- [18] Timmers, P. "Business models for electronic markets", *Electronic Markets - International Journal of Electronic Commerce and Business Media*, Vol. 8, No. 2, pp3-8, 1998.
- [19] Timmers, P., *Electronic Commerce Strategies and Models for Business-to-Business Trading*, John Wiley Publish, 1999.
- [20] Vasilopoulou, K., N. Pouloudi, S. Patronidou, A. Poulymenakou, "Business models: A Proposed Framework," *In Proceedings of the e-Business and e-Work Annual Conference*, Prague, Czech Republic, pp.1003-1009, 2002.
- [21] Vuksic, V., M. Stemberger and J. Jaklic, "Simulation Modeling towards e-business models development," *J. of Simulation*, Vol.2, No. 2, 2002.
- [22] Whitely, D., *E-Commerce - Strategy, Technologies and Application*, McGraw-Hill, 2000.
- [23] Wurman, P. "Dynamic Pricing in the Virtual Marketplace", *IEEE Internet Computing*, Vol.5, pp.36-42, March/April 2001.

Evolution of Service Model and Architecture of ASP Business

Kangrae Cho, Hojoon Choi

Service Development Laboratory, KT, 17, Woomyeon-dong, Seocho-gu, Seoul, 137-792, Korea
{broncho,chopchop}@kt.co.kr

ABSTRACT

ASP has been a buzzword among enterprise IT area since mid 1990s. Following after the hype cycle of new technologies like others, many market analysts say that ASP technology is now in a plateau of productivity sequence through trough of disillusionment sequence over past 5 years. In the Republic of Korea, MIC (Ministry of Information and Communication) and KT began ASP service called bizmeka[®] to help companies adopt the newest information technologies on KT's information superhighway since 2000. Like other ASP players in other countries, there were similar obstacles and problems that are difficult to solve. But with the introduction of new concept of service model and architecture, the difficulties of managing business improved. In this paper, we will discuss about how ASP business of Korea started, managed and made evolution of service model and architecture to meet the various demands of enterprises over past 4 years.

Keywords: ASP, Outsourcing, Enterprise Computing, Web Services

1. OVERVIEW

Today's companies are depending heavily on the newest enterprise software, cutting edge hardware and broadband network to correspond with the rapidly changing market environment. Information technologies that are selected in contemporary corporations give much benefit to them. On the wile, it requires much enterprise resources such as software, hardware and network and operation staff for being beneficial to them. In this point of view, renting an enterprise information technology is much more profitable rather than possessing the whole software, hardware and operation staff considering today's company's situation that must utilize effectively its manpower and money.

Especially in SME (Small Medium Enterprise)'s viewpoint, the needs of renting an enterprise information technology is much bigger than that of a large enterprise. In enterprise information technology market, ASP (Application Service Provider) based service which is capable of constructing enterprise information system rapidly and economically attracted attention of companies from mid 1990's along the boom of the internet. With the advent of ASP business model, ASP sector also won many IT vendors and IT market analysts to its side [2].

Meanwhile, Korea had a different situation compared to other countries with the rapid growth of Internet subscriber over xDSL and cable modem. Most government and public offices and companies have been able to connect to the Internet with low-priced subscription fee and people have come to bear in mind the digital economy and culture gradually. Particularly, to strengthen the domestic and international competitive power of SME in Korea, MIC and KT began an ASP support program. Under the policy of KT's n new business model excavation promotion, bizmeka has been successful in ASP area with the growing revenue and

increasing subscriber since 2000, satisfying customer's needs through evolution of service model and architecture of ASP service platform.

2. MARKET TREND & FORECAST

ASP industry received the spotlight by its rational business model with the Internet boom and Dot-com bubble from 1997 to 2000 by the press. But recently, the descent of IT economy and the collapse of Dot-com bubble burst resulted in merge and acquisition between ASP players and retirement of incompetent ASP players. On the while, the expectancy of ASP market settlement is growing bigger with the overcoming of the hype and immaturity of the new technology and ASP market.

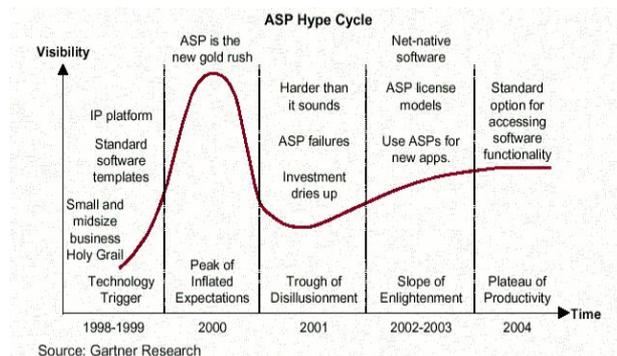


Fig 1. ASP Hype Cycle by Gartner Group

As Figure 1 shows, Gartner expects that ASP technology will be in a plateau of productivity sequence through trough of disillusionment sequence by 2004 following after the hype cycle model of emerging technologies. IDC also expects that the number of ASP player will drop to 389 on the year 2005 from 800 on the year 2000 with the yearly decreasing rate of 13% by severe competition between ASP players, excessive initial investment cost, uncertain business plan and difficulties

in financing. [7]

IT market analysis firms like IDC, Gartner, Ovum are issuing research reports which are being optimistic about current and future ASP market. IDC predicts that the global ASP market size will be 19.9 billion dollars until 2006. Dividing ASP market by region, as Table 1 shows, U.S will occupy 58% of the whole market by 2006 with the market size of about 11 billion dollars and EU market size will be reached to 6.7 billion dollars by 2006 growing from 306 million dollars on 2001. In that, Asia-Pacific ASP market will be 1.5 billion dollars size, but we may expect more big figures in this area considering recent brilliant economic growth of China even considering IDC's forecast. [8].

Table 1. ASP market forecast by region
(Unit: Million US\$)

	2001	2002	2003	2004	2005	2006
U.S.	1,329	2,173	3,494	5,495	8,105	11,667
EU	306	749	1,622	3,036	5,122	6,726
Asia	156	280	441	694	1,071	1,572
Sum	1,791	3,202	5,567	9,225	14,928	19,965

2.1 United States

ASP market in United States showed emergence in 1998 by the companies like Corio and USInternetworking and came into expansion stage during 1999~2000 with many participating ISVs. But the market expansion stage soon stepped into declining stage along with collapse of Dot-com and burst of IT bubble. But from the year 2001, ASP market entered into new era with the M&A between ASP companies and new participant from IT related companies.

There is also continual specialization and subdivision in ASP market after Dot-com crisis. At first, most ASP player thought that price-sensitive SME is their best potential customer for their services. However, there are increasing cases of contract of providing ASP services to a certain company's department partly not to the whole departments. For example, Salesforce provides their Sales Force Automation solutions only to marketing and sales department of America Online, Mitsubishi and Fujitsu

Even new trends and diverse approach method to market, ASP customers in United States are still worrying about problems that are hard to solve like security concern when depositing their valuable data to distant location other than their office and integration problem between ASP application and legacy application and so on. so ASP players are striving to solve these problems by adopting new emerging technologies like EAI (Enterprise Application Integration) or Web Services.

Nowadays, there is an apparent tendency of big IT company's entrance into ASP market. Microsoft has recently announced publication of low-priced CRM software package and participated in ASP business. SAP

and Peoplesoft also have plans to strengthen marketing in SME sector and Oracle has launched new organization for managing new ASP business which rents their solutions on the internet escaping from direct sales over CD. IBM also announced that they are building datacenter infra to prepare for ASP business by packaging their software and solutions into ASP-ready products. [8]

2.2 Japan

In the year 1999 and 2000, ASP business was considered as a good business model with great anticipation. In Japan, but after burst of IT bubble and the collapse of Dot-com companies, Japanese ASP industry also gone into recession period until 2002. But with the explosive growth on broadband internet connection service, ASP industry has been paid attention to again with possibility.

The Japanese ASP market stated with HR, finance, tax management software on the while ASP market in U.S. are ignited by cheap rent of expensive ERP solution. Especially in Japan, ASP gets its attention in public and government sector. Japanese government now drive sharable outsourcing IT infra project by the initiative of 3,300 local provinces. [8]

2.3 EU

IDC predicts that European ASP market will grow 85.6% during 2001~2006 and market size will be 67 billion dollars until 2006. The initial ASP players in European market almost come from United States. Most of the m failed to successfully enter into emerging ASP market. But they are preparing to reenter into new market by new solution, services and technologies.

Most distinguishing project related to European ASP market is "ALTERNATIVE" which driven in public and government sector. The objective of "ALTERNATIVE" project is eventually to strengthen competitive power of SME by allowing selectively choosing their needed solution. [8]

2.4 Summary

We can discover some interesting points by reviewing ASP market trends, status and activities in United States, Japan, and Europe.

Firstly, even big and famous ASP players focus their strength on revenue and profit rather than technological advances by world wide IT economy recession.

Secondly, to win over deepening competition between ASP players, they are trying to seek their own distinguishing services and solutions for customers to discriminate them from other ASP players.

Lastly, multinational companies like Microsoft, IBM,

SAP and Oracle are preparing to enter into ASP market. so existing ASP players are now on the strain for the upcoming competition with well-known package solution vendors.

3. ASP BUSINESS IN KOREA

The Republic Of Korea drives internet boom which is begun from late 1990s and now 30 million users are using the internet and 11 million houses are connected to the internet via xDSL, Cable-Modem or Optic Fibrer at the speed of ranging from 2mbps to 100mbps. In addition, most Korean companies utilize corporate homepages for advertisement and customer contact point.

But comparing the rate of IT infra utilization, the utilization rate of households and large corporations are already reached that of advanced countries. On the while, The IT utilization rate of 3 million SMEs which comprises 99% of Korean companies is apparently low because of its high initial construction cost and management cost. So The Government of the Republic of Korea and KT are doing ASP business called bizmeka for SMEs to inspire the IT mind and spreading of the new enterprise information technology based on KT's broadband internet infra and domestic enterprise solution for SMEs from 2001[3].

On the while, the Korean ASP market size is about 100 million dollars in 2002 and will grow 53% annually until 2007. This growth rate is somewhat dependent on Korean economic recovery on domestic sector. Since most solution provided in ASP industry is mainly focused on domestic business not on global business like export and import goods.

It is shown by survey research that over 71% of SMEs in Korea prefers IT system construction by outsourcing, purchasing and ASP not by operating their own staff and develop their own solution. Especially 44% of SMEs chose the outsourcing as the best preferable method of constructing their own IT infra. It can be easily estimated considering SME's weak IT solution development capability and small IT management budget. We can also expect that SME's dependency and preference over ASP and outsourcing model will be heavier as time passes.

3.1 solutions and services of bizmeka

Starting from web based corporate email and groupware services, bizmeka provides currently about 70 kinds of services ranging from general business solutions from a groupware to core enterprise solutions like ERP, SCM and even special sector solution like IP camera surveillance for 51 thousands companies and 110 thousands employees and employers.

Intra-Enterprise solution sector includes ERP, Production SCM, Groupware, Tax and Finance Management, Sales

and Stock Management, Mobile SFA, IT Equipment Rent and IT consulting service. Collaboration solution sector includes Merchandising SCM, Electronic Tax Receipt and Social and Medial Insurance EDI (Electronic Document Interchange).On-Line solution sector includes CRM, internet shopping mall builderand homepage hosting. Vertical industry solution includes Auto shop management, Optician's Store management and Catering Management.

3.2 architecture of bizmeka service platform

It is very difficult and complicated work to ensure integration between diverse solutions and manage multiple modern enterprises' information system without hassle. To support it stably, perfect and reliable AAA (Authentication, Authorization, Accounting), billing, order management, customer management is key issue of running that kind of system. KT has been running bizmeka platform to run above functions since November, 2002 after one year's research and development

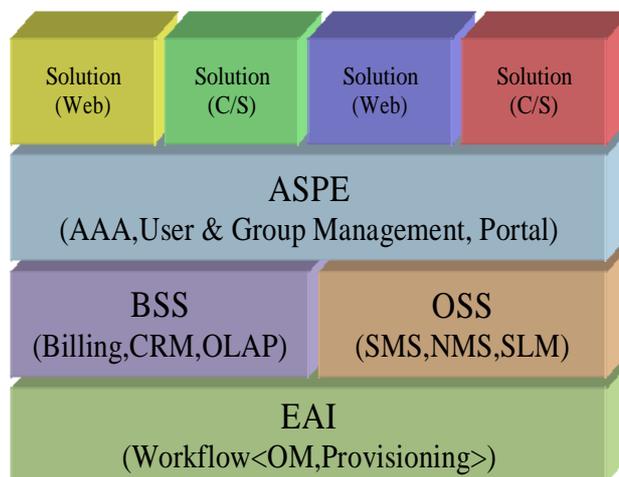


Fig 2. Initial architecture of bizmeka

As Figure 2 shows, ASPE (ASP Enabling) is subsystem that supports AAA of ASP users, organization management and SSO (Single Sign On) over portal. All of them is necessary for using ASP solutions easily and securely. ASPE also has Mobile SMS (Short Message Service), User Community and BBS (Bulletin Board System) which can be used as building blocks of other solutions.

BSS (Business Support System) includes billing management which is crucial for running charged service, ASP user management and OLAP system which gathers and analyze sales related information. Billing system sends a bill to user by sending bill information to existing KT's PSTN (Public Switched Telephone Network) billing system.

OSS (Operation Support System) includes IT resource (Network, Server, Storage, Software) management system to effectively manage massive IT systems. It also

includes SLM (Service Level Management) subsystem which detects failure of services and measures service quality over the network since there are many services and solution based on internet.

EAI system provides reliable message transfer service between servers and message transformation service with orchestration of workflow service. Whole messages are expressed by XML and even heterogeneous systems based on Java or .NET can send and receive messages to each other freely. It also supports reliable, long-running transaction based workflow process like provisioning new services and registering new ASP user after verifying one's credit status.

4. EVOLUTION OF SERVICE MODEL AND ARCHITECTURE

Though bizmeka started with web based solutions on early stage, there was no choice but to accommodate the various demands of companies to meet the customer needs and extend business area. So now bizmeka have general service architecture like EIP (Enterprise Information Portal) and unique and rare kinds of service architecture like VIP (Vertical Industry Portal), MSP (Mobile Service Platform) and APP(Application Publisher Platform). bizmeka also has special kinds of service model such as Remote Hosting and AIP(Application Infrastructure Provider).

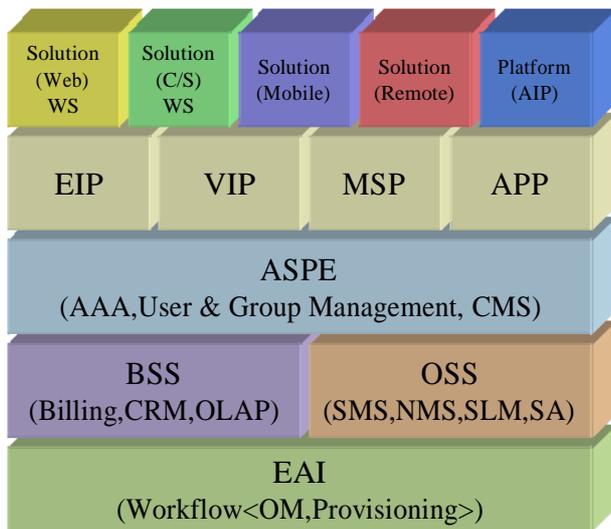


Fig 3. Current architecture of bizmeka

4.1 Portal service platform

EIP is an portal service platform which integrates applications (email, contact list management, bank account management ...) together and provides SSO (Single Sign On), privilege management and personalization to give user the simple way of using the solutions and control authentication and authorization of users on one point. EIP platform includes Multiple Enterprise Support, Site Template Engine, Caching Engine, Content Management Engine, Personalization

Engine, AAA Manager and Service Monitor. With the help of EIP platform, users can aggregate many solutions easily after one successful authentication.

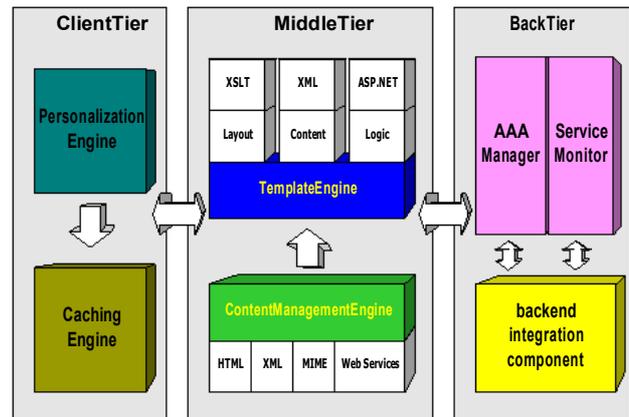


Fig 4. Portal architecture of bizmeka

VIP is designed to help building the vertical industry portal site easily and rapidly. The purpose of EIP is an solution integration but VIP's purpose is to allow building the vertical industry portal in short time using the site template and site management functions in the VIP platform. VIP shares the same portal infrastructure with EIP as shown in Fig. 4. But VIP platform has additional components like site build wizard, handy and easy site management interface and many web interface decorating functions.

4.2 Mobile service platform

MSP is recently added to meet the growing needs of mobile applications and devices like mobile phone, smart phone, PDA and industry specific embedded mobile devices. This platform includes mobile security support system including mobile PKI and data encryption and decryption, contents conversion system supporting HTML, WML, cHTML, WAP and application & data sync system. With the wide acceptance of mobile technologies and devices, the usage of MSP is expected to grow as time goes on.

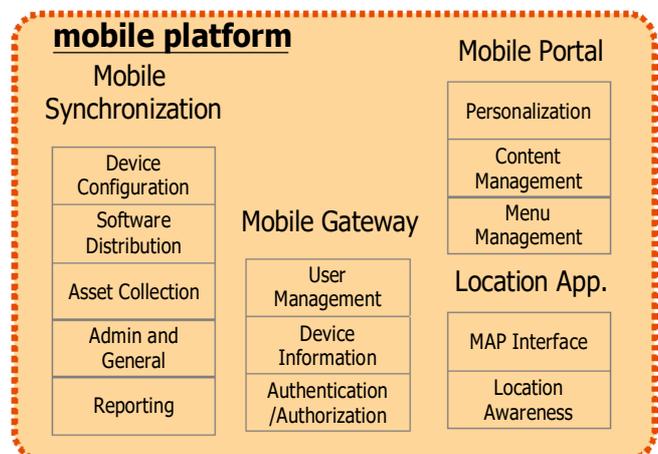


Fig 5. Mobile service platform of bizmeka

4.3 Application publisher platform

APP is a service platform that provides whole or part of platform services through web services to ease the migration of ISV (Independent Solution Vendor)'s solution which already have their customers. In this case, it is really hard to integrate user information and user interface seamlessly to bizmeka. So Application Publisher helps ISVs integrate their solutions to the bizmeka platform loosely and transparently to their existing user.

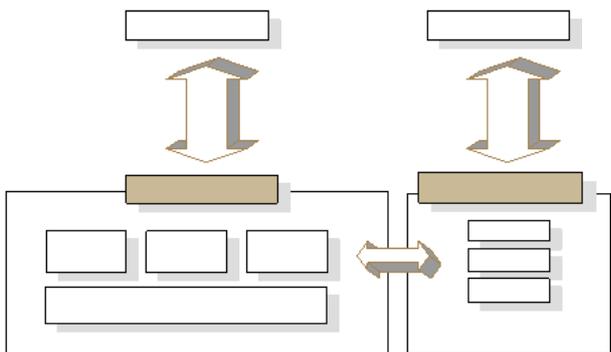


Fig 6. Application publisher Platform of bizmeka

4.4 New kinds of service model

Remote Hosting is a kind of remote management service model for solutions like ERP and SCM which is deployed on customer's site. Operation staff of bizmeka always manage and monitoring the remote applications, servers 24/7. Nowadays the concept of Remote Hosting is extending to the remote network management including virus wall appliances.

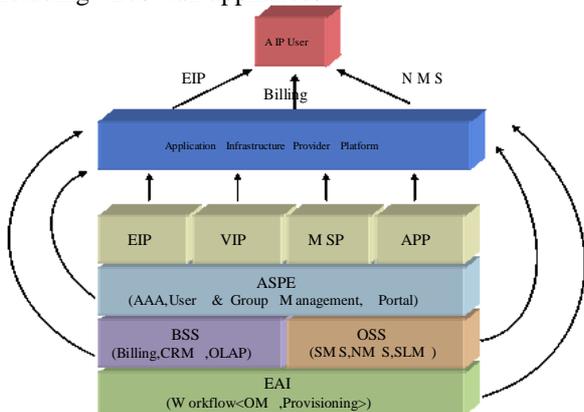


Fig 7. AIP platform usage scenario

AIP is an ASP service platform rent service model that allows independent brand and marketing and admit doing their own ASP business regardless of bizmeka. AIP customers can use the whole platform services or part of them including ASPE (ASP Enabling), BSS (Business Support System), OSS (Operation Support System), EAI (Enterprise Application Integration) subsystem.

5. CONCLUSION

For the past 4 years, bizmeka has been evolved from

simple web application hosting provider to today's ASP accepting various demand of customers and developing experimental service model and architecture. But the trend in ASP business is still changing rapidly and bizmeka needs to excavate new service model and its supporting architecture.

KT also strives to solve fundamental problems of ASP business. Like an unreliable service response time due to jittering internet which is inevitable in current best-effort TCP/IP network. But that can be overcome with successful transition to NGN (Next Generation Network) after 2007. Insecure communication problem is solved by KT's VPN (Virtual Private Network) service which can be bundled with any solutions in bizmeka.

But the most hard-to-solve fundamental problem is anxiety and unrest feelings of customers that their valuable data might be drained to public or opponents. But this problem can be solved by reliable cryptography technologies and technological advance over all IT area in near future.

ACKNOWLEDGEMENT

bizmeka[®] is the official trade mark of KT's ASP business.

REFERENCES

- [1] Jackie Fenn, "The Microsoft System Software Hype Cycle Strikes Again", Gartner, July, 1995
- [2] Che Jong Jin, "KT bizmeka vision and direction of evolution", *The Journal of Korea Information Processing Society*, Vol 10, No 6, 2003
- [3] Korea Information and Telecommunication Industry Association, "Domestic ASP Industry status and demand research", Jul 2003.
- [4] Ben Pring, "2003 ASP Hype Cycle: Hype? What Hype?", Gartner, May 2003
- [5] Meta Group, "ASP models face uncertainty", <http://news.com.com/2009-1017-252761.html>
- [6] Network Briefing Daily, "European ASP market", May 2001.
- [7] Euan Davis, James Eibisch, Lars Schwaner, "ASP Industry Review and Forecast, 2001-2006", IDC, Mar 2002.
- [8] Oh Kwang Seok, "ASP industry trend research", *The Journal of Korea Information Processing Society*, Vol 10, No 6, 2003
- [9] Kevin Newcomb, "The Second Coming of ASPs?", <http://www.aspnews.com>, May 5, 2004
- [10] Lynn Heber, "ASPs Still Alive and Kicking", <http://www.aspnews.com>, Jan 30, 2004
- [11] W3C Working Draft, "Web Services Architecture", <http://www.w3.org/TR/ws-arch/#whatis>, Aug 8, 2003
- [12] Jean Sebastien Mercy, "A Better Understanding of the Enterprise Information Portal Market", http://intranetjournal.com/articles/200110/eip_10_03_01a.html
- [13] IDC, "Worldwide Application Infrastructure Provider Forecast and Analysis, 2000-2005", Jul 200

HTML / HTTP

Integrated Optimization Model of Production Planning and Scheduling for Batch Production

Wenli Shang¹, Yushun Fan²

¹Shenyang Institute of Automation, Chinese Academy of Science, Shenyang, China

²National CIMS Engineering Research Center, Tsinghua University, Beijing, China
shangwl@sia.ac.cn, fan@cims.tsinghua.edu.cn

ABSTRACT

An integrated optimization model of production plan and scheduling is introduced. Function objective is to minimize sum of total setup cost (initial setup cost and related setup cost), stock cost, production cost and overtime cost. Restraint functions include equilibrium stock, production capacity, as well as scheduling restraints (procedure restriction and machine capacity restriction). Feasible plan can be set down with considering scheduling restraints.

Key words: integrated optimization, production plan, scheduling, schedule restraints

1. INSTRUCTION

In machine tool Industry, machine shop can be differentiated into two types: small part with large batch and big part with small batch. The former shop floor produces machine tool axes, cover, tray and special figure part, type of product vary a lot, lot size is large, production cycle is short, plan change frequently, lot size includes monthly batch, semi-monthly batch, weekly batch and daily batch as a result of large lot size delivered and frequently stock. Enterprise plan includes yearly production plan, quarterly production plan and monthly production plan. Manufacturing department in factory deliver monthly product plan to shop floor, and monthly product plan is carried into execution in shop floor. Planner in shop floor set down shorter part production planning, they generally confirm part lot size and turn of delivering parts by experience, thereby result in large work-in-process, infeasible plan and imbalance production. Traditional hierarchical plan decomposed in MRP II often results in infeasible plan in scheduling system as for without considering schedule restraints, new method to establish plan must be adopted.

The kernel goal of plan in batch shop floor is to ascertain part lot size during each period and minimize sum of total setup cost and stock cost. For batch production, literature[1~4] proposed integrated optimization method for production planning and scheduling. Literature [1] put forward integrated model of production planning and scheduling in shop floor, decompose planning alternatively for instances with setup time and without setup time in Job-shop, but the method can only convergence at local optimization results. Literature [2] set up an integrated model of production planning and scheduling in shop floor, schedule restrictions as procedure restraint and machine capacity restraint are taken into account, and its Lagrange relaxation-based solution approach is described, but the method is hard to solve large scale problem for the complexity of the arithmetic to relax

schedule sub-problem. Literature [3] give a heuristic method aiming at complex assembly product manufacturing, optimization lot size is made certain by grouping orders and schedule by searching primary meshwork routes, and function objective is to minimize sum of setup time, stock cycle and total delivery cycle, but lot size in multi-periods is not discussed. In literature [4], an integrated general model of lot-sizing and scheduling is described for multi-stage, multi-resources flexible flow-line, function objective is to minimize sum of total setup cost, stock cost and overtime cost, but production cost and schedule restraints are not considered.

According as merit and demerit of above-mentioned integrated model for production planning and scheduling in batch manufacturing, an integrated optimization model of production planning and scheduling for batch production is recommended, function objective is to minimize sum of total setup time, stock cost, production cost and overtime cost, schedule restrictions include stock equilibrium, procedure restraints, production capacity and machine capacity. The model is non-linear, mixed, integer programming. Genetic arithmetic is an overall optimization searching method with strong adaptability and better extended capacity, its convergence can be amended by linking with heuristic and other searching method, and it can be used to solve complex non-linear problem^[5]. A mixed, heuristic genetic arithmetic will be used to search results of above-mentioned model in future research.

2. INTEGRATED OPTIMIZATION OF PRODUCTION PLANNING AND SCHEDULING FOR BATCH PRODUCTION

In machining workshop of batch production in machine tool industry, type of product varies from several to many, parts processed monthly change constantly with much kinds. Machining workshop has multiplicate equipments, and per equipment has one or more machines. For variable market, shop floor production

planning must be dynamic, real-time and agile. In this paper, an integrated optimization model of production planning and scheduling is described in the context of axes workshop section in shenyang primo machine tool company, which belongs to medium batch production with much kinds. It decomposed workshop sections' monthly part production planning into weekly cell part production planning, and each part's weekly lot size is fixed on.

Axes workshop section belongs to small part with large batch production, includes four production lines: shaft, principal axis, spindle and caudal core, parts are processed separately in each production line, several machine tools are in overtime. Working procedure length, part complexity, cell equipments' capacity, production balance and assemblage of whole product are main factors considered while to set down planning. Scheduling factors include rough material, tools, equipments and staff.

Cell weekly production planning is lot-sizing problem. Lot-sizing planning is to allot batch and process time of process items collection, so as to meet definite capability target. Considering within finite plan span in shop floor, N kinds of parts are processed on M machine tools, and each part is processed on machine tool in same order. Setup time of part j on machine k is correlated to type of former part i processed. Plan span is separated into T period, and exterior requirement d_{jt} for each part in period t is known. Objective of lot-sizing planning is to arrange item lot size in each period, and to minimize sum of total setup cost, stock cost and production cost during whole plan period on the condition of meeting restrictions.

Supposing: 1) Lot size must meet exterior requirements during each period; 2) During each period, a part must be processed without interrupted, other part must be machined after the part; 3) Overtime is permitted, but excess cost brought must be considered.

Integrated optimization model of production planning and scheduling meeting above conditions is as follows:

$$J = \min \sum_{j=1}^N \sum_{k=1}^M \sum_{t=1}^T SC_{0jk} Z_{jt} + \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^M \sum_{t=1}^T SC_{ijk} Y_{ijt} + \sum_{j=1}^N \sum_{t=1}^T (p_{jt} X_{jt} + h_{jt} I_{jt}) + \sum_{k=1}^M \sum_{t=1}^T oc_k O_{kt} \tag{1}$$

Subject to:

$$I_{j,t-1} + X_{jt} - d_{jt} = I_{jt} \quad \forall j \in N, \forall t \in T \tag{2}$$

$$\sum_{j=1}^N \sum_{k=1}^M p_{jk} X_{jt} + \sum_{j=1}^N \sum_{k=1}^M st_{0jk} Z_{jt} + \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^M st_{ijk} Y_{ijt} \leq C_{kt} + Q_{kt}, \forall t \in T \tag{3}$$

$$Y_{ijt} B - X_{jt} \geq 0 \quad \forall i \in \{0, N\}, \forall j \in N, \forall t \in T \tag{4}$$

$$\sum_{j=1}^N Z_{jt} = 1 \quad \forall t \in T \tag{5}$$

$$X_{jt}, I_{jt}, O_{kt} \geq 0 \quad \forall j \in N, \forall t \in T, k \in M \tag{6}$$

$$Y_{ijt} Z_{jt} \in \{0, 1\}, \forall i \in N, \forall j \in N, \forall t \in T \tag{7}$$

Known parameters: d_{jt} —independent requirements of item i in period t; p_{jt} —unit production cost of item j in period t; p_{jk} —machining time of item j on machine k; st_{ijk} —setup time from item i to item j on machine k, if $i=0$, denote initial setup time of item j on machine k; SC_{ijk} —setup cost from item i to item j on machine k, if $i=0$, denote initial setup cost of item j on machine k; h_{jt} —unit stock cost of item j during period j to j+1; C_{kt} —usable throughput of machine k; OC_k —overtime cost of machine k during unit time; B—a very large plus.

Decision-making variables: I_{jt} —stocks of item j on period t; O_{kt} —overtime of machine k on period t; X_{jt} —production quantity of item j on period t; Z_{jt} —setup variables, if $Z_{jt} = 1$, denote setup course to process item j at the end of period t or at the beginning of period t+1, other instance $Z_{jt} = 0$; Y_{ijt} —setup variables, if $Y_{ijt} = 1, i \neq j$, denote setup course to process item j after item i processed, other instance $Y_{ijt} = 0$.

Objective function (1) is to minimize sum of total setup cost (initial setup cost + related setup cost), stock cost and overtime cost during whole plan period. Restriction (2) is essential stock balance equation. Restriction (3) is inequality of capacity restraints. Restriction (4) denotes setup can only happen while production quantity is plus and border upon items are different. Restriction (5) denotes that each item has only one initial setup. Restriction (6) denotes that decision-making variables are not minus. Restriction (7) means that setup variables are 0 or 1. But above-mentioned model not includes following schedule restrictions:

- a. Procedure restraints: each item has several working procedure with definite order.
- b. Capacity restraints: a machine can only process one item on any time.

To describe these schedule restraints, such parameters are defined: J_{jt} —an item batch task (j,t), namely item j processed on period t, its production quantity is X_{jt} , $\forall j \in N, \forall t \in T$; J_t —N task collection must be

completed during period t ,
 $J_t = \{J_{1t}, J_{2t}, \dots, J_{Nt} \mid \forall t \in T\}$; J —all the task
 collection, $J = \bigcup_{t=1}^T J_t$; S_{jt} — setup variables,
 $S_{jt} = \begin{cases} 1, & X_{jt} > 0 \\ 0, & X_{jt} = 0 \end{cases}$; O_{jkt} — operation of job J_{jt} on
 machine k ; t_{jkt} — time begin to operate O_{jkt} ; OM_k —
 collection of all the operations on machine k ,
 $OM_k = \{o_{jmt} \mid m = k, j \in N, t \in T\}$; OP_j —
 collection of all the operations of item j ;
 $OP_j = \{o_{jkt} \mid j = j, k \in M, t \in T\}$; p_{jkt}^u — process
 time of operation O_{jkt} of unit item j ; τ_{jkt} — setup
 time of operation O_{jkt} of unit item j ; GN —collection of
 all the operations with supposed incept and accomplish
 operations; GA — versus collection of working
 procedure composed of all the operations with orders;
 GE_k — versus collection of working procedure
 composed of all the operations on machine k .

In this way, above schedule restraints can be described
 as follow:

$$t_{jkt} - t_{j'kt} - \tau_{jkt} X_{jkt} - \tau_{j'kt} S_{jkt} \geq 0, \quad (8)$$

$$\forall (o_{jkt}, o_{j'kt}) \in GA$$

$$t_{jkt} - t_{j'kt} - \tau_{jkt} X_{jkt} - \tau_{j'kt} S_{jkt} \geq 0 \quad (9)$$

$$\forall (o_{jkt}, o_{j'kt}) \in GE_k$$

$$t_{jkt} \geq 0 \quad o_{jkt} \in GN \quad (10)$$

$$S_{jt} \in \{0, 1\}, \forall j \in N, t \in T \quad (11)$$

Restraint (8) denotes turns between working procedures
 of the same item; restraint (9) denotes turns between
 items processed on the same machine tool; restraint (10)
 shows decision-making variables are not minus;
 restraint (11) expresses that setup variables are 0 or 1.

Integrated optimization model of production planning
 and scheduling with considering lot-sizing and
 scheduling restraints as follow:

$$J = \min \sum_{j=1}^N \sum_{k=1}^M \sum_{t=1}^T sc_{0jk} Z_{jt} + \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^M \sum_{t=1}^T sc_{ijk} Y_{ijt} \quad (1)$$

$$+ \sum_{j=1}^N \sum_{t=1}^T (p_{jt} X_{jt} + h_{jt} I_{jt}) + \sum_{k=1}^M \sum_{t=1}^T oc_k O_{kt}$$

Subject to: formula (2) ~ (11).

3. CONCLUSIONS

Motivated by the aim of improving the feasible plan
 level of short plan period in a real case, an integrated
 optimization model of production planning and
 scheduling has been proposed in weekly item batch
 confirming.

The function objective is to minimize sum of total setup
 cost (initial setup cost + related setup cost), stock cost
 and overtime cost of item within whole planning period.
 To ascertain lot size, detailed schedule restraints are
 discussed, so as to get hold of a feasible production
 planning.

As for schedule restrictions, procedure restraints and
 capacity restraints are debated. Procedure restraints are
 described as each item has several working procedure
 with definite order, and capacity restraints are put
 forward by way of a machine can only process an item
 on any time. A mixed, heuristic genetic arithmetic will
 be used to search the results of above-mentioned model
 in next research work.

ACKNOWLEDGEMENT

This work was supported by the Chinese Hi-Tech R&D
 program under the Grant (2002AA412020).

REFERENCES

- [1] Lasserre J B. An integrated model for job-shop
 planning and scheduling[J]. Management Science, 1992,
 38(8): 1201~1211.
- [2] Xiong R, Chen H S, Hu B S. An integration model
 for production planning and job shop scheduling and its
 lagrangian relaxation-based solution approach [J].
 Journal of Xidian University, 1996, 23(4): 510~516.
- [3] Anwar M F, Nagi R. Integrated lot-sizing and
 scheduling for just-in-time production of complex
 assemblies with finite set-up[J]. Internal Journal of
 Production Research, 1997, 35(5): 1447~1470.
- [4] Xiong H Y, He Y. An integrated model jointing
 lot-sizing and scheduling for flexible flow-line and its
 genetic heuristic algorithm[J]. Journal of changsha
 railway university, 2001, 19(1): 51~55.
- [5] Fleming P J, Purshouse R C. Evolutionary
 algorithms in control systems engineering: a survey [J].
 Control Engineering Practice, 2002, 10(11): 1223~1241

Mapping from BPMN-Formed Business Processes to XPDL Business Processes

Jung, Moon-young, Hak Soo Kim, Myung Hyun Jo
Kyung Hyun Tak, Hyun Seok Cha, Jin Hyun Son

Department of Computer Science and Engineering, Hanyang University, Korea
{myjeong, hagsoo, mhjo, khtak, hscha, jhson}@cse.hanyang.ac.kr

ABSTRACT

In the business process management, many business process execution languages such as XPDL, BPML, BPEL4WS have been specified with different origins and goals. Most of all, XPDL proposed by WfMC has been widely used in the related applications, especially workflows whose concepts are currently interchangeable with those of business processes. On the other hand, Business Process Modeling Notation (BPMN) driven by BPMI has recently been specified as a standardized graphical notation for a business process. We can therefore commonly design and analyze various business processes using the design tools to support BPMN. Notice that a BPMN-formed business process should be converted to its semantically equivalent business process languages such as XPDL which can consequently be executed by business process engines. In this regard, we propose a transformation mechanism from BPMN-formed business processes to corresponding XPDL processes.

Keywords: BPMN, XPDL, business process

1. INTRODUCTION

With an increasing number of business applications that automate business processes, many enterprises have recently devoted considerable attention to business process integration. The primary goal of business process integration is inter-operating the information flows between IT organizations that have used a number of various terms to describe how components can be connected together to build complex business processes. For instance, if our system tries to connect a partner during operating a business process, that is composed of tasks such as catalogue request, order, order processing, and dispatching, we must be required business process integration technologies. It is necessarily required many costs for reducing the difference between methodologies of each vendor. For merging business processes, which are located in distributed environment, there is a standard interface that provides flexibility or interoperability between business processes. To streamline business operations, it is important to define a standard business process language in the first place that could be adapted in complex business circumstances. It is a same reason that the computing resource is composed by the programming language such as C or Java. However, a business process language is not relied on a physical object such as operating systems or devices, but it must be designed suitably in a logical meaning of a business environment. Especially, it should be defined a mechanism that one internal partner connects dynamically the other external partners in distributed environments. In order to meet these requirements, researches of the business process language have been broadly going to two trends.

The first of these trends is a research in business process modeling language. It is lately remarkable by Business

Process Modeling Notation (BPMN) Working Group. Although a business analyst doesn't know any information about an internal mechanism of a business process execution, it allows him to design a business process using a business process modeling notation. Examples of other notations or methodologies that were reviewed are UML Activity Diagram, UML EDOC Business Processes, IDEF, ebXML BPSS, Activity-Decision Flow (ADF) Diagram, RosettaNet, LOVeM, and Event-Process Chains (EPCs). BPMN has more functionality and extensibility than other business process modeling languages. So, we select BPMN for mapping from a business process modeling language to a business process execution language.

The second of these trends is a research in business process execution language. It enables a system to understand all of the information which makes our business runs. Representatively, Business Process Modeling Language (BPML) is developed by the Business Process Management Initiative (BPMI), Business Process Execution Language for Web services (BPEL4WS) is written by developers from BEA Systems, IBM, and Microsoft, and XML Process Definition Language is announced by the Workflow Management Coalition (WfMC). And they are XML-based business process execution languages. To minimize a modification of existing workflow applications using XPDL, we select XPDL among a number of business process languages.

Although there is a significant difference between two trends, both in motivating business process modeling and in their features, they have a strong common basis. Thus, the convergence of two major trends is creating a rapidly growing demand for a new breed of software that facilities automation of business processes both between enterprises and within an enterprise. However,

the difference between business processes represents a partial meaning difference, but it raises enormously the risk. In this paper, we describe how the mapping mechanism reduces a meaning difference between BPMN and XPDL. Figure 1 illustrates the mapping structure of this paper; in the future we will plan to study mapping mechanisms from BPMN-formed business processes to BPEL4WS and BPML business processes.

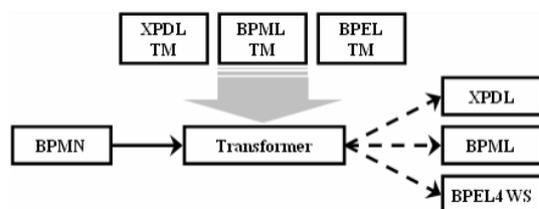


Figure 1. Transformation structure of BPMN

The remainder of the paper is organized as follows. Section 2 presents an overview of related work. Section 3 describes Mapping from BPMN to XPDL. Section 4 contains concluding remarks.

2. RELATED WORK

There are several researches in mapping from business process modeling notation to business process execution language [1,3,4,5,6]. This Section describes mapping mechanisms which have already been proposed.

[3] describes the mapping from BPMN to XPDL. It shows that BPMN notations can be transformed to XPDL straightforwardly, because of similar business process structures within them. However, there are BPMN notations which can not be mapped to XPDL, straightforwardly. Therefore mapping is applicable to only 10 BPMN notations. So, we will show the mechanism which can be applicable to other BPMN notations, either, in this paper.

[1] describes the mapping from BPMN to BPEL4WS. It is accomplished by analyzing cases which BPMN business process can be expressed. It also describes all elements and attributes of BPMN mapping to BPEL4WS elements. So, the number of pages of [1] is more than 50. It may be complete, but it is too complex to be proved.

[4] and [5] describe the mapping from UML Diagram to XPDL. [4,5] shows business process can be expressed using UML Diagram, and be mapped to XPDL. [4] describes it using Use Case Diagram, Statechart Diagram, and Activity Diagram with stereotypes, which is defined extendedly, such as <<Route>>, <<No>>, <<Tool>>, <<Subflow>>, <<Loop>>. [5] describes it using only Activity Diagram with stereotypes, which is defined extendedly, such as <<Business Document>>.

[6] describes the mapping from UML Diagram to BPEL4WS. It shows business process can be expressed using Class Diagram and Activity Diagram with

stereotypes, which is defined extendedly, such as <<Process>>, <<Activity>>.

As shown in above, many researches select UML Diagrams as a notation for business process modeling more actively than BPMN. However, if we use UML Diagram to model business process, and if we want to transform the business process to a business process language, we will need to extend the stereotypes which are dependent on transformed languages.

Besides, there are several researches in expression power of business process modeling notations or business process execution languages [7,8,9,10,11,12]. (e.g., workflow patterns, Petri-net)

Like this, mapping mechanisms, which the existing papers propose, are straightforward, case by case analyzed, or system dependent formed. If a mechanism is straightforward, it will be not complete, since a difference of a notation and a language. If a mechanism is case by case analyzed, it will be too complex to be mapped, since it has to be analyzed in all cases. And if a mechanism is language dependent formed, it will be not appropriate for a standard mechanism. Therefore this paper proposes a mapping mechanism from BPMN to XPDL, which is not only simple, complete, but also system interchangeable.

3. MAPPING FROM BPMN TO XPDL

Both BPMN and XPDL are conceived of as a directed graph structure. So Mapping from BPMN to XPDL may be described straightforwardly. e.g., A Task and a Sub-Process of BPMN are transformed to an Atomic Activity and a Subflow Activity of XPDL respectively, and a Sequence Flow of BPMN is transformed to a Transition of XPDL [3]. But because BPMN is designed to model and to manage business processes and XPDL is designed to execute business processes, we need to consider some differences between BPMN and XPDL. Firstly BPMN has several elements that are not for executing business processes but only for modeling or managing business processes. These BPMN elements do not need to map to XPDL elements. Secondly, BPMN has several elements that are for executing business processes but do not transformed to XPDL elements straightforwardly. The business processes, which are represented using BPMN including these elements, have to be transformed to the same meaning XPDL business processes.

This Section describes mapping from BPMN to XPDL as follows. In section 3.1 we extract a BPMN element set, which consists of the BPMN elements that have to map to XPDL elements. In section 3.2 we look around the BPMN-formed business processes as a structural point of view and compare these with XPDL business processes. In section 3.3 we describe the mapping from BPMN to XPDL straightforwardly and in section 3.4 we

propose the transformation mechanism for BPMN elements which are not mapped to XPD elements.

3.1 Analysis on BPMN Elements

Business processes are designed by business analysts using BPMN notations, these BPMN-formed business processes are transformed to XPD processes for executing business processes, and the executing business processes are managed by business process managers in BPMN format. That is, the BPMN consists of elements for business process design, elements for business process execution, and elements for business process management, so several BPMN elements, which are not for business process execution, don't need to be transformed to XPD elements.

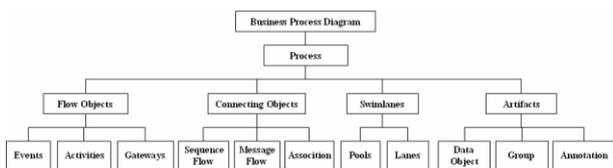


Figure 2. Major elements defined within BPMN

Figure 2 shows the major elements of BPMN and those categories. In the Business Process Diagram, the graphical objects (Flow Objects, Connecting Objects, Swimlanes, and Artifacts) define Processes (business processes); Flow Objects define the behavior of Processes, Swimlanes group other elements and assign roles to them, Artifacts provide additional information about the Processes, and Connecting Objects represent the order of Flow Objects or association between each element. But Swimlanes, Artifacts, and some of Connecting Objects (Message Flow and Association) are not related to the business process execution. First, a role, which is represented by Swimlanes, is a semantic element which means the role of doing the work (e.g. Seller, Buyer, Client, and etc.). But the element, which we need to know for executing business process, is not a role, but a performer. Second, Artifacts and Association are not directly related to the flow of process. Third, Message Flow shows the flow of messages between two entities, but the actual message is not transmitted by Message Flow but by applications or web services which is bound by Flow Objects. Therefore we don't need to transform these BPMN elements to XPD elements. Moreover, each BPMN elements have their attributes, and some of the attributes (i.e. Name, Author, Language, CreationDate, ModificationDate, GraphicElements, Status, Categories, InputSets, Input, OutputSets, Output, BoundaryVisible) are not related to business process execution. And there is no need to transform these attributes to XPD elements, either. Figure 3 shows the BPMN elements, except the elements which are described above. That is, these elements are related to business process execution and have to be transformed to XPD elements.

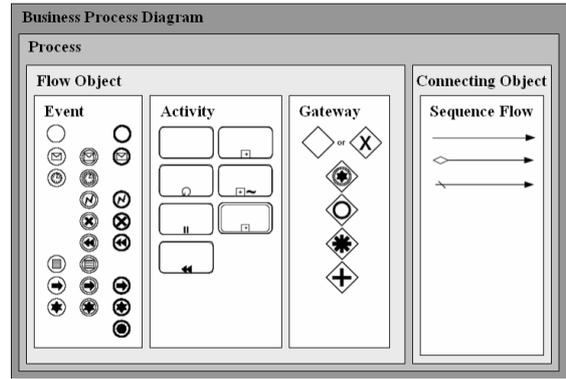


Figure 3. BPMN elements which have to be transformed to XPD elements

3.2 Structural Mapping

A business process is a behavioral flow which is structured by performers, services, and data. In a business process, works, which is performed by performers using services and data, are defined orderly.

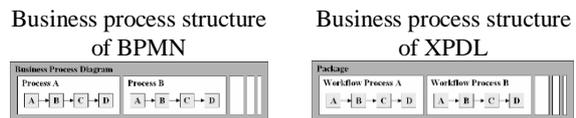


Figure 4. Business process structures of BPMN and XPD

BPMN and XPD represent business processes similarly, as shown in Figure 4. Processes, which are defined by BPMN elements, are structured by Flow Objects and Connecting Objects in directed graph format. Also, Workflow Processes, which are defined by XPD, are structured by Activities and Transitions in directed graph format. Like this, Business Process Diagram, Processes, Flow Objects, and Sequence Flows in BPMN are mapped to Package, Workflow Process, Activity, and Transition in XPD respectively.

Table 1. Structural mapping

BPMN element	XPD element
Business Process Diagram	Package
Process	Workflow Process
Flow Object	Activity
Connecting Object	Transition

And performers, services, and data which are the elements of business processes are represented Performer, Implementation, Property in BPMN, and Performer, Tool, Data Field in XPD.

Table 2. Business process elements mapping

BPMN element	XPD element
Performer (within Activity)	Performer (within Activity)
Implementation (within Activity)	Tool (within Activity)
Property (within Process and Activity)	Data Field (within Workflow Process)

3.3 Simple Mapping

In Section 3.2, we described that BPMN Flow Objects and Connecting Objects are mapped to XPD L Activities and Transitions. But BPMN Flow Objects and Connecting Objects are classified with Events, Activities, Gateways, and Sequence Flows, and these are classified more detailed according to the value of type related attribute. This Section describes mapping from these detailed BPMN elements to XPD L elements which are transformed straightforwardly.

3.3.1 Events

BPMN Event acts as an event, which occurs, or reacts against another event. (e.g., cancel of an order, modification of an order, and handling these events) And it is classified according to its position (Start, Intermediate, End) and the type of Trigger (Message, Timer, Error, Cancel, Compensation, Rule, Link, Complex, Terminate).

Table 3. BPMN Event mapping to XPD L

BPMN element	XPD L element
None Start Event	Route Activity
None End Event	Route Activity
Message Start Event	Route Activity and Formal Parameters of Workflow Process
Message End Event	Atomic Activity and Route Activity
Acyclic Timer Start Event	Route Activity and No Implementation Atomic Activity including Deadline
Acyclic Timer Intermediate Event (Normal Flow)	No Implementation Atomic Activity including Deadline
Acyclic Timer Intermediate Event (Exception Flow)	Deadline of Atomic Activity

In XPD L, there are some elements which it functions as a BPMN Event. But because XPD L is on the basis of data-based control, many of them are not supported. Therefore, we describe these Events, which is not transformed to XPD L yet, in Section 3.4.

3.3.2 Activities

BPMN Activity is work that is performed within a business process. Activity is classified into Sub-Process and Task according to the subject of the work, and these are classified more detailed according to the method that it works. And we can transform it into XPD L Activity as shown in Table 4 and Table 5.

Table 4. BPMN Sub-Process mapping to XPD L

BPMN element	XPD L element
Embedded Sub-Process	Block Activity
Independent Sub-Process	Subflow Activity
Reference Sub-Process	Block Activity or Subflow Activity equivalent to the Activity referenced

Table 5. BPMN Task mapping to XPD L

BPMN element	XPD L element
Service Task	Atomic Activity
Receive Task	Atomic Activity
Send Task	Atomic Activity
User Task	Atomic Activity including Performer
Script Task	Atomic Activity including Extended Attribute which can contain the script
Manual Task	Manual Mode Atomic Activity
Reference Task	Activity equivalent to the Activity referenced

We didn't describe the Mapping from Ad-Hoc Sub-Process to XPD L yet. But because it is a kind of complex mapping, we describe it in Section 3.4.

3.3.3 Gateways

BPMN Gateway controls the flow of both diverging and converging Sequence Flow. And the types of Gateway are classified according to split/merge and the function. In XPD L, Route Activity implement split or join transitions, and it works as BPMN Gateways according to Transition Restriction element within Route Activity.

Table 6. BPMN Gateway mapping to XPD L

BPMN element	XPD L element
Exclusive Decision (XOR) -Data-Based	XOR Split Route Activity
Exclusive Merge (XOR) -Data-Based	XOR Join Route Activity
Inclusive Decision (OR)	AND Split Route Activity
Inclusive Merge (OR)	AND Join Route Activity
Parallel Fork (AND)	AND Split Route Activity
Parallel Join (AND)	AND Join Route Activity
Complex Decision / Merge	Combination with several Route Activities and Transitions

We didn't describe the mapping from Event-Based Exclusive Decision to XPD L yet. But because it is a kind of complex mapping, we describe it in Section 3.4.

3.3.4 Sequence Flow

BPMN Sequence Flow shows the order of Flow Objects. And the types of Sequence Flow are classified according to the value of Condition attribute. In XPD L, as described in Section 3.3, Transition implements like this. And the type of Transition is classified according to Condition element, either.

Table 7. BPMN Sequence Flow mapping to XPD L

BPMN element	XPD L element
Sequence Flow	Transition
Conditional Sequence Flow	Transition including CONDITION Type Condition
Default Flow	Transition including OTHERWISE Type Condition

3.4 Complex Mapping

Complex mapping is the mapping for BPMN elements (complex elements) which XPD L didn't have as the same function. That is, if we recompose a business

process, which is composed of BPMN elements including the complex elements, to a business process, which is composed of BPMN elements except complex elements, we can transform the business process to XPD. It is the same as if "A = B" and "B = C", then "A = C". This paper proposes following three mechanisms for these transitive mapping.

Mechanism 1. Loop

The Loop means that it works more than one time. We can apply this mechanism to the complex element, if the element has the semantic of "it works repeatedly" or "it works again".



Figure 5. Loop transformation mechanism

Mechanism 2. Discriminator

The Discriminator means that it completes only a work of all other works which are implemented concurrently. We can apply this mechanism to the complex element, if the element has the semantic of "it finishes B or C".

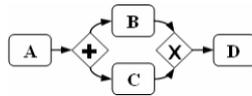


Figure 6. Discriminator transformation mechanism

Mechanism 3. Serialization

The Serialization means that it transforms something, which is serialized implicitly, to another thing serialized explicitly. We can apply this mechanism to the complex element, if the element has the semantic of "it works in the order of a basis".



Figure 7. Serialization transformation mechanism

Firstly, Cyclic Timer Event, Standard Loop Activity, Sequential MultiInstance Loop Activity, and StartQuantity attribute of Activity apply to the Loop; Cyclic Timer Event represents the work which is implemented in a cycle. Standard Loop and Sequential MultiInstance Loop Activity represents the work which is implemented repeatedly according to the Condition of Activity. StartQuantity attribute of Activity represents the work which waits until the number of token is satisfied. And the Figure 8 shows the application of the Loop transformation mechanism to these BPMN elements.

Secondly, Message Intermediate Event (Exception Flow), Error Event, Multiple Event, Terminate Event, and Event-Based Exclusive Decision apply to the

Discriminator; Both Message Intermediate Event (Exception Flow) and Error Event mean that whether it flows normal flow or exception flow. Multiple Event means that if one of the Event, which is assigned to itself, occurs, it will be triggered. Terminate Event means that it interrupts all the works, which is working in the process, and end the process at once instead of normal end. Event-Based Exclusive Decision means that only one of the Events, which are in order, can be triggered. And the Figure 9 shows the application of the Discriminator transformation mechanism to these BPMN elements.

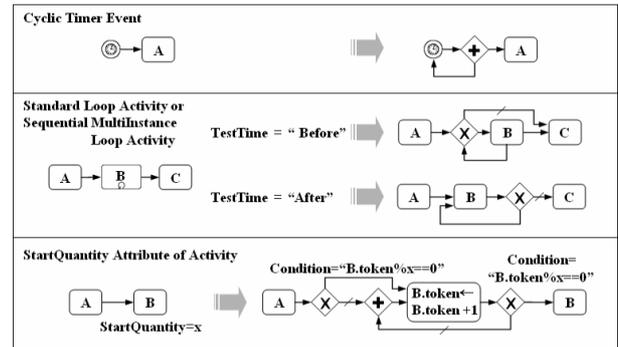


Figure 8. Loop transformation class

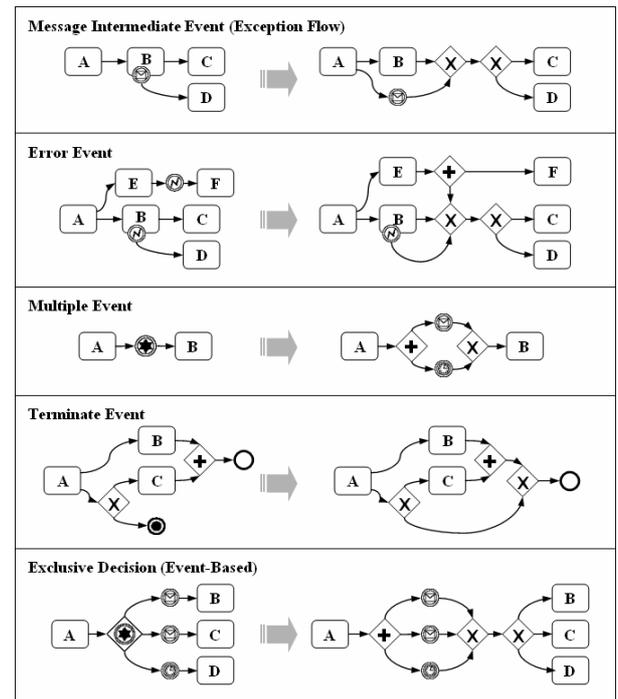


Figure 9. Discriminator transformation class

Thirdly, Link Event and Parallel MultiInstance Loop Activity apply to the Serialization; Link Event represents the work which is to be connected to another Link Event which has same LinkId. Parallel MultiInstance Loop Activity means that it works same work in parallel. And Figure 10 shows the application of the Serialization transformation mechanism to these BPMN elements.

Lastly, Ad-Hoc Sub-Process is the Process which the order of the works and the end of the process is decided by performers at execution time. So it works beginning with the work, which is selected by a performer, repeatedly until the CompletionCondition of the Ad-Hoc Process is satisfied. Therefore Ad-Hoc Process applies to both Loop and Discriminator. Figure 11 shows the application of the transformation mechanisms to Ad-Hoc Sub-Process.

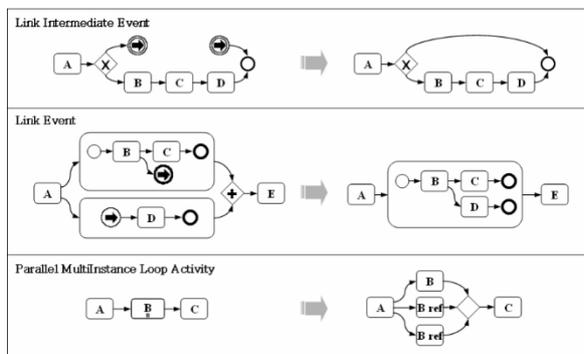


Figure 10. Serialization transformation class

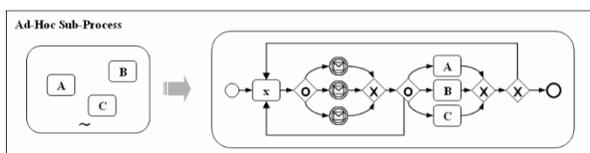


Figure 11. Transformation of Ad-Hoc Sub-Process

4. CONCLUSIONS

This paper focused on a problem which has arisen from a meaning difference between BPMN and XPD. We have analyzed elements that are mapped from BPMN to XPD, and then have proposed three transitive mapping mechanisms (loop, discriminator and serialization).

While the existing mapping mechanisms only selected elements that can be directly mapped, we discovered all of elements that are able to be mapped with a notation's transformation without a meaning loss. These methodologies transform a complex BPMN element into several kinds of atomic and simple BPMN elements. That is, because a reconstructed business process is composed of elements, which are adapted to mapping to XPD, we can transform the BPMN- Formed Business Processes to XPD Processes.

We are going to develop a module with this mapping mechanism. We will also plan to study mapping mechanisms from BPMN to other business process execution languages such as BPEL4WS and BPML.

ACKNOWLEDGEMENT

This work was supported by Korea Research Foundation Grant (KRF-2004-003-D00327).

This work was supported by grant No. R08-2003-000-10464-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

This work was supported by the Ministry of Information & Communications, Korea, under the Information Technology Research Center (ITRC) Support Program.

REFERENCES

- [1] BPMI.org., Business Process Modeling Notation (BPMN), Version 1.0 - May 3, 2004.
- [2] WfMC., Workflow Process Definition Language -- XML Process Definition Language (XPDL), Document Number WfMC-TC-1025, Document Status - 1.0 Final Draft. October 25, 2002.
- [3] Stephen A. White., "XPDL AND BPMN", Workflow Handbook 2003.
- [4] P. Jiang, Q. Mair, and J. Newman., "Using UML to Design Distributed Collaboration Workflows: from UML to XPDL", *The Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, 2003.
- [5] R. Eshuis, P. Brimont, E. Dubois, B. Grégoire, and S. Ramel., "Animating ebXML Transactions with a Workflow Engine", *The Eleventh International Conference on Cooperative Information Systems (CoopIS 2003)*, Catania, Sicily, Italy, 2003.
- [6] T. Gardner., "UML Modelling of Automated Business Processes with a Mapping BPEL4WS", *The 17th European Conference on Object-Oriented Programming (ECOOP)*, Germany, 2003.
- [7] Workflow Patterns Home Page, Available from <http://www.workflowpatterns.com>
- [8] Stephen A. White., "Process Modeling Notations and Workflow Patterns", Workflow Handbook 2004.
- [9] W.M.P. van der Alast., "Patterns and XPDL: A Critical Evaluation of the XML Process Definition Language", 2002.
- [10] P. Wohed, W.M.P. van der Aalst, M. Dumas, and A.H.M. ter Hofstede., "Pattern Based Analysis of BPEL4WS", QUT Technical report, FIT-TR-2002- 04, Queensland University of Technology, Brisbane, 2002.
- [11] W.M.P. van der Aalst, M.Dumas, A.H.M. ter Hofstede, and P. Wohed., "Pattern Based Analysis of BPML (and WSCI)", QUT Technical report, FIT-TR-2002-04, Queensland University of Technology, Brisbane, 2002.
- [12] Robert Shapiro., "A Technical Comparison of XPDL, BPML and BPEL4WS", 2002.
- [13] David Hollingsworth., "The Workflow Reference Model: 10 Years On", Workflow Handbook 2004.

Methodological Approach for Choosing Portfolio of E-Business Models

Ayoob Mohamadian¹, Fereidoun Ghasemzadeh², Emad Farazmand¹

¹School of Management, Tehran University, Tehran, Iran

² School of Management and Economics, Sharif University of Technology, Tehran, Iran
mohmadian@ut.ac.ir

ABSTRACT

As seen in the wide range of researches, the lack of a proper and profitable business model is an important factor in companies' failure and in the use of Internet and e-business. In this paper we want to determine the best (or portfolio of) e-business models for each organization by which they can use Internet and information technologies in the optimal manner in their business. To do so, we will consider strategies, organizational and environmental conditions of each organization and determine the best e-business models in the base of these conditions. So two steps appraisal process be introduced, that can improve the earlier frameworks and remove their disadvantages and also it can assist managers in the choice phase of decision-making.

Keywords: electronic business models, decision making process, organizational characteristics

1. INTRODUCTION

The Internet is challenging the economic, social, and technological foundations of the old economy. In essence a revolution is taking place, entrepreneurs developing new models for business, the economy, and government. A business model is a method of doing business by which a company can generate revenue to sustain itself. The model spells out how the company adds value, which customers are willing to pay for, in terms of the goods and/or services the company produces in the course of its operations (Turban, 2002, p.8). Many companies are starting to use the Internet to communicate with both, their customers and suppliers, creating new digital electronic commerce networks that bypass traditional distribution channels. The Internet, the web, and intranets can help companies achieve new levels of competitiveness and efficiency, but one of the management challenges that they raise is finding a successful Internet business model (Laubom, K. et al, 2001, pp.251-252).

Nowadays from all sides –from investors, stock market analysts, employees, customers, suppliers and competitors –have pressured to senior managers of 'Old Economy' companies to migrate from marketplace to marketpace (Weil et al, 2001), in other side developing e-business models for conducting e-business is not simply about the adoption of new technologies. It also concerns changes in work practices, in customer/supplier relationships, in the way products are delivered to consumers, in marketing practices and changes in staff skills needed to support e-business. Accordingly, e-business models signify new opportunities for re-organizing the way businesses are currently practiced (Vassilopoulos and et al., 2001). It can be said that the companies' first action in moving toward e-business is the choice of the best e-business model because in e-business models all the changes that are mentioned are regarded and explained. But nobody has introduced a

complete methodology for choosing e-business models up to now. Also the evaluation of all the e-business models is very difficult .so this paper attempt to introduce a methodology for choosing the proper e-business models for organizations by introducing the factors that effect on the portfolio of e-business models.

In proposed methodology shortcomings of prior works is removed and all phases of Simon's decision-making process (intelligence, design, choice) is covered so systematic decision making for choosing e-business models is presented. In order to introduce the mentioned methodology, the result of Hayes and Tjan works are addressed.

The paper is organized as follows. The second section presents the e-business model concept and an overview of current understanding of e-business models in the literature. Section three expresses the Simon's decision-making process. Section four explains e-business models classifications as they are reported in the current literature and Then Tjan's model is introduced. In Section 5, Methodological approach for evaluating e-business models in real world is proposed and finally suggestion and conclusions is expressed.

2. E-BUSINESS MODEL CONCEPT

The benefits of living in a fast changing information society would become evident if effective e-business practices are implemented, offering companies exposure and access to global markets and consumer access to customized, high quality services. Innovative and effective e-business models facilitate such practices (Vassilopoulos and et al., 2001).

The term *Business Model* rose to prominence In the 1990s with the advent of IT-centered businesses. The rise of the term is closely related to the emergence and diffusion of commercial activities on the Internet.

Internet start-ups used the term to differentiate themselves from the incumbents and to explain their competitive position. The term appeared first in 1970s in computer science journals, after 1995 also in popular business and computer magazines like *Business Week* or *Wired* before it gained access into peer-reviewed journals focusing on the emerging field of e-commerce and e-business. Today the term is frequently used also in other management journals like *Harvard Business Review* (Stahler, 2001). Accelerating growth of IT and ICT has raised the interest for changing traditional business models or developing new business models that better exploit the opportunities offered by technological innovations. That is why within the last few years, the discussions about business models and the impact of the Internet on them have become more topical (Pateli et al., 2002). However, there is still no common understanding of how a business model is defined, how to develop business models, which are the principal business models, how to transform business models, etc (Alt and Zimmerman, 2001). In other words, there is no consolidated knowledge considering various aspects of business models.

A research made by the Institute of Strategic Change of Accenture (Linder & Cantrell, 2000) concluded that:

“Developing a sound business model matters for making money. However, business models wear out, and firms must alter them in order to remain viable. The better managers know their business model, the more better they can manage patterns of change”. That reveals that understanding business models is a prerequisite for organizations that wish to change their business to remain competitive. A number of cases in which the understanding and use of business model is essential:

1. Understanding the key elements and mechanisms in a specific business domain and their relationships (Osterwalder & Pigneur, 2002).

2. Communicating and sharing the understanding of a business model among business or technology stakeholders (Gordijn & Akkermans, 2001c).

3. Experimenting with innovative business concepts to determine if current business models can be easily adapted to new concepts (Eriksson & Penker, 2000), as well as to assess the viability of new business initiatives (Weill & Vitale, 2001).

4. Specifying valid requirements for the Information Systems that support the business model (Eriksson & Penker, 2000)

5. Identifying options for changing and improving the current business model (Eriksson & Penker, 2000), thus facilitating change (Osterwalder & Pigneur, 2002).

Table 1. Business Model Definitions

Researchers	Definition
Magretta (2002)	A story that explains how an enterprise works.
Petrovic et al. (2001) Auer & Follack (2002)	A description of the logic of a “business system” for creating value that lies behind the actual processes.
Jutla, Bodorik, Wang, (1999)	The business model determines processes and transactions. (i.e. business process- retail [external, internal], procurement, transaction- buy, payment registration etc.)
Applegate (2001)	A description of a complex business that enables study of its structure, the relationships among structural elements, and how it will respond to the real world.
Timmers (1998)	Architecture for the product, service and information flows, including a description of the various business actors and their roles; a description of the potential benefits for the various business actors; and descriptions of sources of revenues.
Osterwalder & Pigneur (2002)	A description of the value a company offers to one or several segments of customers and the architecture of the firm and its network of partners for creating, marketing and delivering this value and relationship capital, in order to generate profitable and sustainable revenues streams.
Weill & Vitale (2001)	A description of the roles and relationships among a firm’s consumers, customers, allies and suppliers that identifies the major flows of product, information, and money, and the major benefits to participants
Tapscott et al. (2000)	A business model is about the invention of new value propositions that transform the rules of competition, and mobilize people and resources to unprecedented levels of performance.

Several attempts have been made so far to define e-business models. Table 1 presents the widely cited definition provided by Timmers (1998) as well as some other definitions suggested more recently in the literature. These definitions range from generic (Magretta, 2002; Petrovic et al., 2001) to more concrete ones (Timmers, 1998; Weill & Vitale, 2001; Osterwalder & Pigneur, 2002). Thus, we can find definitions that explain what the purpose of a business model is, while other definitions focus on specifying its primary elements, and possibly their interrelationships.

Furthermore some researchers perceive the Business Model as a purely business concept that explains the logic of making business for a firm (Timmers, 1998; Linder & Cantrell, 2000; Petrovic, 2001; Rappa, 2001), while some others consider it as a link between strategy, business processes, and information systems (Nilsson et al., 1999; Osterwalder & Pigneur, 2002). The difference between these two interpretations of Business Models concerns the relationship of Business Model with the concepts of Strategy, Business Processes, and Technology. While in the first

interpretation the three concepts are included in the description of Business Model, the second interpretation considers them as inter-linked components set in different levels of a pyramid construct. In this paper, a business model is considered as the conceptual and architectural implementation (blueprint) of a business strategy and represents the foundation for the implementation of business processes and information systems.

3. DECISION MAKING PROCESS

Decision-making is regarded as one of the most important activities carried out by an organization (Simon, 1960). Simon has distinguished two extremes regarding the structuredness of decision problems. At one end of the spectrum are well-structured problems that are repetitive and routine and for which standard models have been developed. Simon calls these programmed problems, example of such problem is weekly scheduling of employees. At the other end of the spectrum are unstructured problems, also called nonprogrammed problems, which are novel and non-recurrent such as evaluating an electronic commerce initiative. Semistructured problems fall between the two extremes (Turban et al, 2001,p.43). The decision to engage in electronic business can be seen as non-programmed as new strategies and new technologies are required (Hayes et al., 2003).

Systematic decision-making process according to Simon, involves three major phases: intelligence, design, and choice. Simon's model is the most concise, and yet completes characterization of rational decision-making. In the intelligence phase the reality is examined and the problem is identified and defined. Importance of this phase is that impossible solutions are eliminated which saves in time and money. In the design phase, potential alternative solutions are identified and criteria are set for evaluation of the alternative courses of action. The choice phase includes selection of a proposed solution to the model (not to the problem it represents). (Turban et al., 2001,p41)

Choice is the critical act of decision-making. The choice phase is the one in which actual decision is made and where the commitment to follow a certain course of action is made. The boundary between the design and choice phases is often unclear because certain activities can be performed during both the design and choice phases and because one can return frequently from choice activities to design activities. The choice phase includes search, evaluation, and recommendation of an appropriate solution to the model.

Little research has been conducted on determining appropriate ebusiness models and strategies despite its importance to organizational decision-makers. Finnegan and Golden (1995) propose that the adoption

of electronic business applications is based on two factors; market receptivity and the probability of competitor adoption. Coltman et al. (2001) highlight the importance of the organizational environment to decisions regarding electronic business. Hayes and Finnegan (2003) have attempted to help decision makers with the intelligence phase in relation to assessing the potential of e-business models. Their research identified a set of organizational factors, or organizational prerequisites, that are critical for the successful assessing ebusiness models. Decision makers can use this list of organizational prerequisites to help focus decision-making activity at the intelligence phase.

This paper aims to utilize the prerequisites concept to develop a methodological approach to help managers focus on all phases of decision-making (intelligence, design, choice) by excluding electronic business models that are not compatible with prevailing organizational, environmental/supply chain conditions in first level of evaluation and choosing of appropriate ebusiness model by considering other factors that introduced by Tjan (2001) in second level of evaluation. In fact this proposed approach is extended of Hayes's model.

4. E-BUSINESS MODELS CLASSIFICATION

A great deal of research has been directed towards classifying business models and grouping them into specific categories. The business models belonging to the same category usually share some common characteristics, such as the same pricing policy or the same customer relationship model Pateli, Giaglis, 2003. The taxonomy frameworks of Business Models that are presented in the literature differentiate based on two factors: a) *Criteria* posed for classifying Business Models, b) *Objects classified*, whether they are entire business initiatives (such as Amazon, eBay, etc), possibly combining multiple business models (Timmers, 1998; Rappa, 2001), or atomic business models that can be incorporated into an ebusiness initiative (Weill & Vitale, 2001). The most common sets of criteria are:

- *Interaction Pattern and Value Chain Integration* (Timmers, 1998),
- *Functional Integration and Degree of Innovation* (Timmers, 1999),
- *Core Activities and Price –Value Balance* (Linder & Cantrell, 2000),
- *Economic control (both hierarchical and self-organizing) and value integration* (Tapscott et al., 2000),
- *Sourcing: What businesses buy (manufacturing versus operating inputs) against how they buy (systematic versus spot sourcing)* (Kaplan & Sawhney, 2000).

- *Degree of value integration (low to high) and control (self-organizing to hierarchical) in the supply chain (Ticoll et al., 1998)*

The existence of multiple sets of criteria suggests that there are no established and commonly agreed to criteria for classifying business models, but some criteria, such as Value Integration, may be used in various taxonomy frameworks. Moreover, most taxonomy frameworks seem to be narrowly defined for Internet e-Business models. A notable exception is found in Tapscott et al. (2000), who introduce a rather generic taxonomy of Business Models that could apply not only to Internet but also to any type of electronic business featured by a network structure.

In order to construct the proposed model, we have examined classification theories of particular electronic business models based on the work of different researchers specially Hayes et al. (2003) and Tjan (2001). It is clear from such research that particular models are more likely to operate under certain conditions.

Hayes et al. (2003) by combining the work of Ticoll et al. (1998), Timmers (1999), and Kaplan and Sawhney (2000) proposed that electronic business models could be classified according to how they exhibit varying degrees of economic control, value chain integration, functional integration, business innovation and technical innovation. They developed the prerequisites framework for assisting decision makers assess the suitability of electronic business models during the intelligence phase of the decision making process and not in other phases. The basis of their prerequisites model is shown in Table 2. This framework hypothesizes that a particular business model is more likely to succeed in a particular industrial sector when the characteristics of the sector match the conditions required for the model. The framework has the potential to help decision makers by providing a

method of excluding from consideration those electronic business models that are unsuitable for given prevailing organizational and environmental characteristics.

In this paper we extend this framework to consider all phases of Simon's decision-making process (intelligence, design, choice) by utilizing Tjan's model as second level evaluation because in this model financial and human aspects and other factors that influence the viability of particular electronic business models have been identified.

Tjan (2001) adopted a business project portfolio applications approach to create an Internet portfolio-planning matrix. However, instead of trading off industry growth and market position, here the strategy is based on *company fit*, which can be either low or high, and the *project's viability*, which can also be low or high. Together these create an *Internet portfolio map (matrix)*. A business model's viability can be assessed on a scale of 1 to 100, by four criteria: market-value potential, time to positive cash flow, personnel requirements, and funding requirements. Then, an average score (simple average) for each metric is computed. For *fit*, the following criteria are used: alignment with core capabilities, alignment with other company initiatives, fit with organizational structure, fit with company's culture and values, and ease of technical implementation. Then each of business models is assessed on a qualitative scale of high, medium, low. The various e-business models are then mapped on the *Internet portfolio matrix*, based on the average scores for viability and fit. The Internet matrix is divided into four cells, as shown in Figure 1. If both *viability* and *fit* are low, the project is killed. If both are high, then the project is adopted. If *fit* is high, but *viability* is low, the project is sent to redesign. Finally, if the *fit* is low but the *viability* is high, the project may be sold or spun off.

Table 2. Characteristics of E-Business Models (Hayes et al., 2003)

Business model	Economic Control	Functional Integration	Supply chain Integration	Innovation	Sourcing
E-shop	Low	Low	Low	Low	Systematic
E-mall	Low	Medium	Low	Medium	Systematic
E-procurement	Medium	Medium	Medium	Medium	Systematic
E-auction	Low	Medium	Medium	Medium	Spot/systematic
Information brokerage	Low	Low	Low	High	Not applicable
Trust services	Low	Low	Low	Medium	Not applicable
Third party marketplace	High	High	High	High	Spot/systematic
E-hubs	High	High	High	High	Spot/systematic
Virtual communities	Medium	Medium	High	High	Not applicable
Value-chain integrators	High	High	High	High	Not applicable
Value-chain service providers	Low	Medium	Medium	High	Not applicable
Collaboration platforms	Low	High	High	High	Not applicable

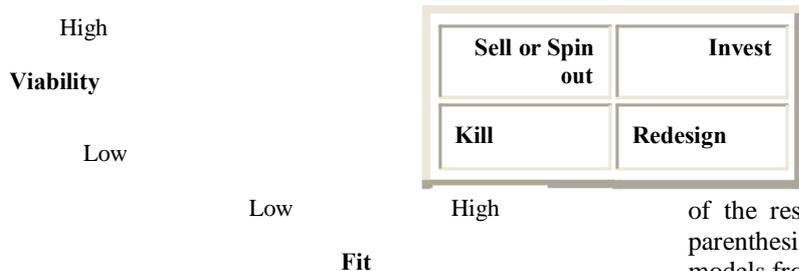


Figure1. Internet Portfolio Map (Tjan, 2001)

5. USING THE PROPOSED METHODOLOGICAL APPROACH

This methodological approach is used to change the current business model toward proper e-business model and it will be useful for the companies that are doing there business traditionally and at the same time want to experience e-business and use of internet in their operations. To do so, in this section a framework is proposed that helps organizational decision makers to select the proper e-business model by a two step appraisal process. In the primary step of this appraisal process, at first those models, which have less chance to be selected, would be identified and ignored for the next step. Doing so results in reduction in cost and time because there will be little number options to be appraised in detail. You can find more detail about this step in the next section. This approach has been improved in comparison with previous models and has removed their disadvantages. Also it has covered all of the Simon’s decision-making phases (intelligence, design and choice) and has introduced a systematic procedure to make decision about business model choice. Another specification of this framework is introduction of new criteria that can be useful in e-business model selection.

5.1 Primary appraisal: deletion of unsuitable business models

This step of appraisal covers intelligence and design phases of decision-making process. The aim of this step is to exclude those electronic business models that are not suitable given prevailing organizational and environmental characteristics. So in the next step there will be the little number of business models that should be appraised more thoroughly and therefore it results saving in time and cost of choosing proper e-business model. Using Hayes model is proposed method for identification of unsuitable business models, because by using this model it is possible to determine organizational and environmental condition by measuring economic control, value of supply chain integration, value of functional integration, degree of innovation and sourcing. In table 3 we can see the rating scales were completed by senior managers within the food company known to the authors. The calculated scores and associated rating for the company

are shown in Table 3. Low, medium and high rating are determined as being 0–34%, 35–69%, and 70–100% (respectively) of the available marks. The ratings in Table 3 were then compared with the characteristics of e-business models previously shown in Table 2. An extract of the results is shown in Table 4. The ratings in parenthesis are those required for individual e-business models from Table 2. The five models in the table are unsuitable due to the differences between the company scores and the model characteristics. So in this step we ignore them and appraise the remained 7 business models.

5.2 Secondary appraisal: selection of suitable business models

In this step the remained models from the previous step that are potentially suitable for the company are appraised more accurately, regarding Tjan model’s criteria. For this purpose, at first quantitative criteria of *viability* for each of the e-business models is assessed in the 0 to 100 scale (table 5) and then qualitative criteria of *fit* assessed in the base of intellectual scales (low, medium, high) (table 6). Then, by regarding the average of obtained scores, the position of regarded business models is determined on the Internet Portfolio Matrix (fig 2). As we can see, regarding implemented appraisal, the company should employ portfolio of e-shop, e-procurement and information brokerage models.

6. CONCLUSION

Today there are too many pressures on companies to use Internet and Information Technology in their business activities and go forward to e-business. But almost they don’t know how to use this new knowledge and technologies. In this paper a methodological approach introduced that can be decision-makers’ guideline in selection of suitable business model. In this framework the previous models disadvantages have been removed and also it has covered all of the Simon’s decision-making phases (intelligence, design and choice) and has introduced a systematic procedure to make decision about business model choice. Also by performing a two-step appraisal process, it ignores the business models that have less probability to be selected, in the first step and cause to reduction in cost and time for choosing suitable business model. The Proposed approach is conceptual, and has not been empirically tested at this point. Our findings from the pre-test of this model are positive. However, further testing is required.

ACKNOWLEDGEMENT

We are grateful to Jeremy Hayes for his intellectual support to this research.

Table 3. Summary data for the company

	Economic Control	Supply Chain Integration	Functional Integration	Innovation	Sourcing
Score	87/150	76/153	93/151	93/125	-
Percentage	58%	50%	61%	74%	-
Rating	Medium	Medium	Medium	High	Spot/Systematic

Table 4. Comparing the framework with company ratings

Business model	Economic Control	Supply chain Integration	Functional Integration	Innovation	Sourcing
Virtual Communities	Medium (Medium)	Medium (High)	Medium (Medium)	High (High)	Systematic (Not applicable)
E-hubs	Medium (High)	Medium (High)	Medium (High)	High (High)	Systematic (Spot/systematic)
Third Party Marketplace	Medium (High)	Medium (High)	Medium (High)	High (High)	Systematic (Spot/systematic)
Collaboration Platform	Medium (Low)	Medium (High)	Medium (High)	High (High)	Systematic (Not applicable)
VC Integrators	Medium (High)	Medium (High)	Medium (High)	High (High)	Systematic (Not applicable)

Table 5. Assessing viability of e-business models

Business model	Alignment with core capabilities	Alignment with other company initiatives	Fit with organizational structure	Funding requirement	Ease of technical implementation	Average
1.E-shop	H	M	H	H	H	H
2.E-mall	M	M	H	L	M	M
3.E-procurement	H	H	H	H	H	H
4.E-auction	M	M	L	H	M	M
5.Information brokerage	M	H	H	H	H	H
6.Trust services	L	M	L	L	L	L
7.VC service providers	L	L	L	L	M	L

Table 6. Assessing fit

Business model	Market value potential	Time to positive cash flow	Personnel requirement	Funding requirement	Average
1.E-shop	80	90	85	95	85
2.E-mall	65	45	70	60	60
3.E-procurement	80	95	80	65	80
4.E-auction	55	60	40	65	55
5.Information brokerage	45	85	95	95	80
6.Trust services	5	10	15	5	10
7.VC service providers	15	25	25	15	20

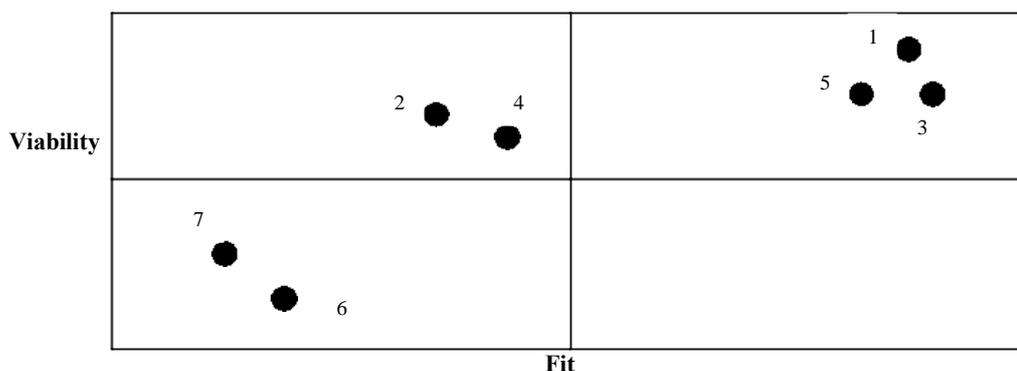


Figure 2. Portfolio of e-business models

REFERENCES

- [1]. Alt, R., Zimmerman, H-D., Preface—Introduction to special section—business models. *Electronic Markets* 11 (1), 3–9, 2001
- [2]. Applegate, L. M., E-business Models: Making sense of the Internet business landscape, In G. Dickson, W. Gary, and G. DeSanctis (Eds.), *Information Technology and the future enterprise: New models for managers*, Upper Saddle River, N.J.: Prentice Hall, 2001
- [3]. Auer, C. and Follack, M., Using Action Research for Gaining Competitive Advantage out of the Internet's Impact on Existing Business Models, In the Proceedings of the 15th Bled Electronic Commerce Conference – eReality: Constructing the eEconomy, Bled, Slovenia, June 17 –19, pp. 767-784., 2002
- [4]. Coltman, T., Devinney, T.M., Latukefu, A., Midgeley, D.F., E-business: Revolution, evolution or hype? *California Management Review* 44 (1), 57–85, 2001
- [5]. Eriksson, H. and Penker, M., *Business Modeling with UML – Business Patterns at Work*, John-Wiley & Sons, New York, 2000
- [6]. Finnegan, P., Golden, W., Electronic commerce: Implementation issues for Irish management. In: Walsh, J. (Ed.), *Proceedings of the Management Research in Ireland Conference*, Cork Ireland, September 12–13. pp. 1–10, 1995
- [7]. Gordijn, J., Akkermans, J.M., Designing and Evaluating E-Business Models, *IEEE Intelligent Systems*, 16(4), 11-17, 2001
- [8]. Hayes, J., Finnegan, P., “Assessing the of potential of e-business models: towards a framework for assisting decision makers”, *European Journal of Operation Research*, 2003
- [9]. Jutla, D., N., Bodorik, P. and Wang, Y., WebEC: A Benchmark for the Cybermediary Business Model in E-Commerce. *IMSA 1999*: 388-392. Landscape, Institute for Strategic Change, Accenture, 1999
- [10]. Kaplan, S., Sawhney, M., 2000. E-hubs: The new B2B marketplaces. *Harvard Business Review* (May–June), 97–103, 2001
- [11]. Laubom, K. and Laubom, J., “Essential of Management Information System”, Prentice-hall, 2001
- [12]. Linder, J.C. and Cantrell, S., *Changing Business Models: Surveying the Institute for Strategic Change*, Accenture, 2000
- [13]. Magretta, J., Why Business Models Matter, *Harvard Business Review*, May, pp. 86-92, 2002
- [14]. Nilsson, A.G., Tolis, C., Nellborn, C., *Perspectives on Business Modelling – Understanding and Changing Organisations*, Springer, Berlin, 1999
- [15]. Osterwalder, A., and Pigneur, Y. *An eBusiness Model Ontology for Modeling eBusiness*, In the Proceedings of the 15th Bled Electronic Commerce Conference – eReality: Constructing the eEconomy, Bled, Slovenia, June 17 –19, 2002, pp. 75-91, 2002
- [16]. Pateli, A., Giaglis, G., ‘A Framework for Understanding and Analysing e-Business Models’, in Proceedings of the 16th Bled Electronic Commerce Conference – eTransformation, Bled, Slovenia, June 9–11 (CD-ROM –Proceedings), 2003
- [17]. Pateli, A., Giaglis, G., “A Domain Area Report on Business Models” White Paper, 2002
- [18]. Petrovic, O., Kittl, C. Teksten, R.D. Developing Business Models for eBusiness, In the Proceedings of the International Conference on Electronic Commerce 2001, Vienna, Austria, October 31 –November 4., 2001
- [19]. Rappa, M., *Business Models on the Web*, 2001, available at http://ecommerce.ncsu.edu/models/models_text.html
- [20]. Simon, H.A., *The New Science of Management Decision*. Prentice Hall, Englewood Cli., NJ, 1960
- [21]. stahler, p., “business models as an unit of analysis for strategizing”, 2002
- [22]. Tapscott, D., Ticoll, D. and Lowy, A., *Digital Capital –Harnessing the Power of Business Webs*, Harvard Business School Press, Boston, 2000
- [23]. Ticoll, D., Lowy, A., Kalakota, R., In: Tapscott, D. (Ed.), *Blueprint to the Digital Economy: Creating Wealth in the Era of e-Business*. McGraw-Hill, New York, 1998
- [24]. Timmers, P., "Business Models for Electronic Markets", *Electronic Markets -International Journal of Electronic Commerce & Business Media*, vol. 8, no. 2, pp. 3-8., 1998
- [25]. Timmers, P., *Electronic Commerce: Strategies and Models for Business-to-Business Trading*. Wiley, 1999
- [26]. Tjan, A. K., “Finally, A Way to Put Your Internet Portfolio in Order,” *Harvard Business Review*,
- [27]. Turban, E., Aronson, J., Bolloju, N., “Decision Support Systems and Intelligent Systems”, Prentice-Hall, Inc., 2001
- [28]. Turban, E., Mc lean, E. and Wetherbe, J. “Information Technology for Management” By John Wiley & Sons, Inc, 2002
- [29]. Turban, E., Lee, J., King, D. and Chung, H.M., *Electronic Commerce: A Managerial Perspective*, Prentice-hall, Upper Saddle River, NJ, 2002
- [30]. Vassilopoulou, K., Pouloudi, A., Patronidou, S. and Poullymenakou, A. “E-Business models: A Proposed Framework”
- [31]. Weill, P., Vitale, M.R., *Place to Space: Migrating to eBusiness Models*, Harvard Business School Press, Boston, 2001.

Rationality Validation of Business Process Model by Simulation Method

Ke Ning¹, Yuliu Chen², David O'Sullivan¹

¹ Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland

² Department of Automation, Tsinghua University, Beijing, 100084, China
Ke.ning@deri.org

ABSTRACT

Business process modeling has been adopted widely. Due to the complexity, it is very hard to validate the rationality of process models. The existing validation methods can only detect structural conflicts in process models. It is also very important to validate those objects related to the processes. This paper presents a rationality validation method which is based on discrete event simulation technology. It can detect three logic mistakes from business process models: structural deadlock, lack of synchronization and objects not matched each other. This method has extended the scope of rationality validation, and also enriches the contents that can be validated.

Keywords: Business Process Models, Rationality Validation, Simulation

1. INTRODUCTION

Business process model is kind of description of business process. By using model, we are able to analyze the performance of process, to guide the implementation of process, to monitor the execution of process, to enable the management of process. Due to the complexity of business process, whether describing AS-IS models or designing TO-BE models, it is possible to make kinds of mistakes about rationality. It is nonsense to analyze the performance of a process based on the wrong models, and implementing a wrong model may bring huge losing to an enterprise. Therefore, more and more attention has been pay to the validation of business process model.

Validation of a business process model is not so easy. If the structure and size are not limited, it will be an NP hard problem to validating a process model [1]. Thus many efforts have been pay to how to improve the efficiency of validation algorithm, hoping to get high efficiency through simplifying model or adding restrict conditions. H. Lin [2] gave a graphic reduction method to validate the mistakes of structural deadlock and lack of synchronization of a process based on his process modeling language. W.M.P. [3] defined a WF-NET based on Petri-Net, then used Petri-Net theories to validate the rationality of the structure of a process. These methods have gain efficiency of multinomial time complexity. But problems raised here: Firstly, there are rigorous conditions for the algorithms, the method of H. Lin can validate only some specified simple structure of process, method based on Petri-Net required that the process model must be a free-choice net. Secondly, these methods can only find some logic problem of process structure, which is only a little scope of process rationality validation. These methods may have theoretical values, but they are not the practical ones for the real business process.

To support validation of real business process, we have gone on a different way. We haven't cared too much about efficiency of algorithm, but have paid more attention to enlarge the scope of contents of validation, and to improve the practicability. We have developed a validation method based on simulation technology in our eIDEF3 business process model. This method can detect three logic mistakes in business process: structural deadlock, lack of synchronization and objects not matched each other. Compared to the existing methods, it is not so effective, but enough for most business process. And most important, it has extended the scope of rationality validation, and also enriches the contents that can be validated.

2. EIDEF3 PROCESS MODELING METHOD

IDEF3 is one of the IDEF series methods. It is used to capture and describe business process. It can support communication and comprehension between domain experts and modelers, and has been used widely in industry. To increase its description ability, support simulation analysis, enlarge model usability, we have extended IDEF3 to eIDEF3 (extended IDEF3), which can describe objects involved in process in a formal way. We have developed a software tool GEM-EASY IDEF3[®] based on this method. Figure 1 is the basic syntactical elements of eIDEF3.

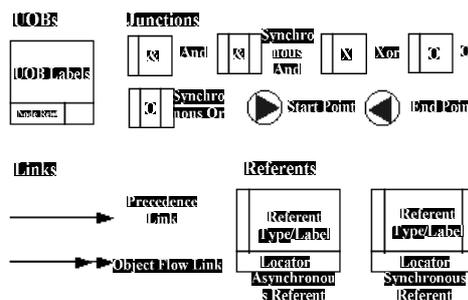
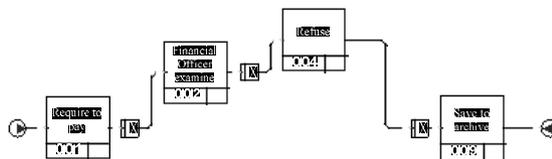
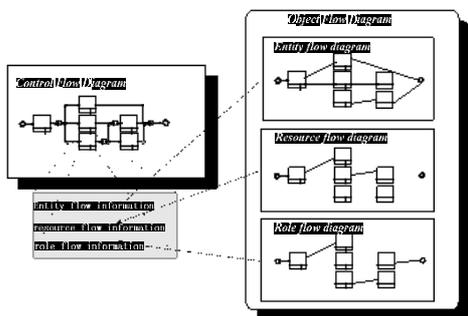


Figure 1 basic syntactical elements of eIDEF3

eIDEF3 provides two type basic flow diagram to describe a process scenario (As shown in Figure 2): One is Control Flow diagram, the other is Object Flow Diagram.



$$\begin{aligned} &\forall n \in N, \\ & \text{if}(n \neq \text{BeginOf}(pis) \wedge n \neq \text{EndOf}(pis)) \\ & \text{then}(n \notin PB \wedge n \notin PE) \end{aligned} \quad (5)$$

It is obviously that this definition of process nationality emphasizes the normally beginning and normally ending of a process. This is consistent with people's intuitive comprehension. More important, this definition is given from process itself, and is not bounded with specified modeling methods. So it is more generic. Moreover, defining process rationality based on PIS has built the foundation to validate it by using simulation method.

4. PROCESS RATIONALITY CRITERION IN EIDEF3 MODEL

Before give the concrete algorithm, we will give the criterion based on some precondition and assumption.

4.1 Basic precondition and assumption

Based on the requirement of eIDEF3 syntactical rules, the following basic precondition must be satisfied.

1) An eIDEF3 model P is an all connected diagram from Start Point to End Point. No isolated nodes exist. That is:

$$\begin{aligned} &\forall n \in P, \\ & \text{InOf}(n) \geq 0 \vee \text{OutOf}(n) \geq 0 \end{aligned} \quad (6)$$

2) Connection rules: The Precedence Links of UOB are single in and single out. The Precedence Links of FanOut Junction are single in and multi out. The Precedence Links of FanIn Junction are multi in and single out. The Precedence Links of Start Point are none in and multi out. The Precedence Links of End Point are multi in and none out. That is:

$$\begin{aligned} &\forall n \in P, \\ & \text{if}(n \in \text{UOB}) \text{then}(\text{InOf}(n) \leq 1 \wedge \text{OutOf}(n) \leq 1) \\ & \text{if}(n \in \text{FanOut_Junction}) \text{then}(\text{InOf}(n) \leq 1) \\ & \text{if}(n \in \text{FanIn_Junction}) \text{then}(\text{OutOf}(n) \leq 1) \\ & \text{if}(n = \text{Start_Point}) \text{then}(\text{InOf}(n) = 0) \\ & \text{if}(n = \text{End_Point}) \text{then}(\text{OutOf}(n) = 0) \end{aligned} \quad (7)$$

Moreover, we give some basic assumption here, which is only for facilitating description of validation problems, and will not influence the applicability of the algorithm:

1) Synchronous and Asynchronous Junctions are not differentiated. Alternate from synchronous to asynchronous, or from asynchronous to synchronous, will not change the Process Rationality.

2) There is no "Or" junction in process models. Based on the semantic of the "Or" junction, we can use a combination of "And" and "Xor" junction to replace it.

3) There is no feedback in a process. We assume that a process is a DAG (Directed Acyclic Graph). If a feedback is needed, we can put the feedback parts into the decomposed scenario, so that we can avoid dealing with it when using the validation algorithm. Then the rationality of the feedback parts can be determined by other methods.

4.2 Criterion of Process Rationality

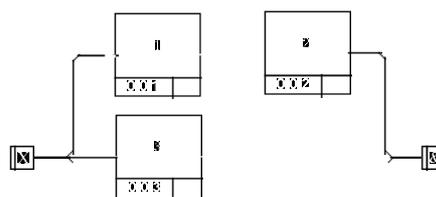
We have provided three criterions to validate process rationality, below are the details:

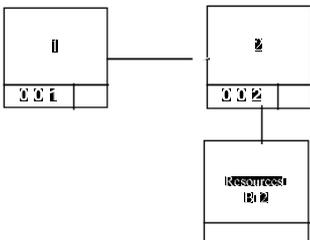
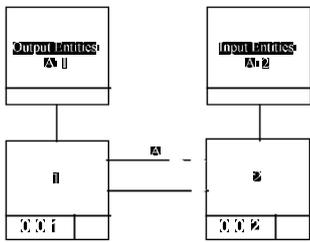
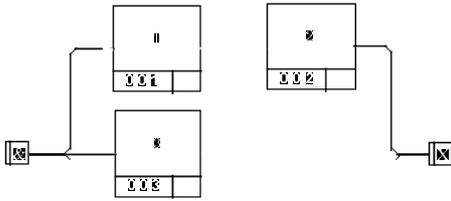
1) Structural Deadlock, SD

If two or more of the FanOut branches of "FanOut Xor" junctions joint in a "FanIn And" junction, SD will happen. That is:

$$\begin{aligned} &\exists n \in \text{FanOut_Xor_Junction}, \\ & \text{OutOf}(n) \geq 2 \wedge \\ & \exists n' \in \text{FanIn_And_Junction}, \\ & n' \in \text{OutPath}(n)_i \wedge n' \in \text{OutPath}(n)_j, \text{ where } i \neq j \\ & \Leftrightarrow P \text{ is SD} \end{aligned} \quad (8)$$

As shown in figure 4, in this situation, the "FanIn And" junction will not be able to be triggered because it can not finish all of its FanIn branch. Thus the process can not execute any more.





```

ELSE
{
    Get a rational pis;
}
4. After getting a rational pis, select randomly one of the rest
FanOut branch from the tailender "FanOut Xor" junction,
repeat step 3, create a new pis;
5. Repeat step 4, until all the "FanOut Xor" junction have
been deal with. Then we get all the PISs of the process. Then
we can conclude that this process is a rational one;
END

```

What have to point out is, this algorithm can find the mistake of rationality, and find the node where the mistake exists. But it can not correct the mistake automatically. You have to correct it manually, and execute the algorithm again, because new change may bring new mistakes in other place. Until no mistakes are found, you get a rational process finally.

5.2 Analysis of the algorithm

1) Completeness

From the step 4 and step 5, we can see that for a rational process, this algorithm can deal with all the fanout branch of the "FanOut Xor" junction. So we can say that this algorithm can explore all PISs of the process. Once there is mistake in the process, the algorithm will come to step 3. Here it can tell where the mistake is. Users can use this message to correct models, and run the algorithm again. Until all mistakes have been corrected, the process will become rational. According to the analysis of last paragraph, all the PISs of a rational process can be explored completely. Therefore, this algorithm is completed. It can explore all PISs of a process, and find all mistakes in it.

2) Complexity

It is easy to conclude that for a rational process, to explore all the PISs, the time consumed will increase exponentially along with the amount of "FanOut Xor" junction:

$$O = \bar{c}_{path}^{C_{FanOut_Xor_Junction}} \quad (14)$$

Here, \bar{c}_{path} is the average amount of fanout branch of the "FanOut Xor" junctions.

3) Usability

As we have mentioned at the beginning of this paper, most existing validation methods have many limitation in their usability. For example, the method of H. Lin can not validate complex models; the method of W.M.P requires the models must be free-choice. Our method has no rigorous limitation on process models, and the contents that can be validated are more abundant and complete.

Compared to those methods that are based on Petri-net, the complexity of our method is much better. For most

enterprises, a business process will not include too many "FanOut Xor" junctions (where decision making is needed). In the Chinese Aviation CIMS Project, we had build lots of business process models, among these models, the most complex model had no more the 10 "FanOut Xor" junctions. For today's computer, it is very easy to deal with such a complexity.

5. CONCLUSION

This paper presents a simulation-based business process rationality validation method. It has extended the scope of rationality validation, and also enriched the contents that can be validated. Although it is not the most efficient one, but for most business processes it is enough. So it is a validation method that is very suited for business process validation. Although this method is created based on eIDEF3, but the concepts about process rationality, process instance sub-graph, are independent of concrete modeling methods. So it is very easy to extend this method to other models, e.g. Petri-net models, workflow models, etc.

ACKNOWLEDGEMENT

The main result of this paper was created during my doctoral research, which was supported by China 863 Program. The publication of this paper was supported by Scientific Fundation of Ireland.

REFERENCES

- [1] Hofstede A.H.M., Orłowska M.E., Rajapakse J. Verification Problems in Conceptual Workflow Specifications. Proceedings of the 15th International Conference on Conceptual Modeling, volume 1157 of Lecture Notes in Computer Science. Cottbus, Germany: Springer-Verlag, pp73-88, 1995.
- [2] H. Lin, Z. Zhao, H. Li, and Z. Chen. A Novel Graph Reduction Algorithm to Identify Structural Conflicts[A]. Proceedings of the Thirty-Fourth Annual Hawaii International Conference on System Science. IEEE Computer Society Press, 2002. 536, 540, 541, 546, 548, 550.
- [3] W.M.P. vander Aslst, Arthur HM, Ter Hofstede. Verification of Workflow Task Structures: A Petri-Net-Based Approach. Information Systems, 2000, 25(1): 43~69.
- [4] NIU Dong, NING Ke, LI Qing, CHEN Yuliu. Method and Tool for IDEF3 Based Process Modeling[J]. Computer Integrated Manufacturing Systems, 2001, 7(12): 30-34.
- [5] NING Ke, NIU Dong, LI Qing, CHEN Yuliu. IDEF3-based Business Process Simulation Modeling[J]. Computer Integrated Manufacturing Systems-CIMS, 2003, 9(5): 351-356.
- [6] Aalst W.M.P. Verification of Workflow Nets. Application and Theory of Petri Nets 1997, volume 1248 of Lecture Notes in Computer Science, Berlin: Springer-Verlag, 1997. 407~425.
- [7] Wasim Sadiq, Maria E.Orłowska. Analyzing Process Models Using Graph Reduction Techniques. Information Systems, 2000, 25(2): 117~134.

Research on Mobile Agent-based E-Commerce System Framework

Wenna Liu¹, Deli Yang²

School of Management, Dalian University of Technology, Dalian 116024, China

¹ wenal@sina.com, ² somdyang@dlut.edu.cn

ABSTRACT

Electronic commerce is one of the most important application areas of mobile agent technology. A secure mobile agent system (SMAS) model is established for e-commerce environment. The SMAS consists of mobile agents and mobile agent server. A security shell module in the mobile agent architecture is responsible for mobile agent protection. Surveillance agent running in the agent server provides the host against attacks from malicious agents. Based on the SMAS, mobile agent-based e-commerce system framework is built up and realized in the end.

Keywords: electronic commerce, SMAS, empirical study, mobile agent

1. INTRODUCTON

Web-based e-commerce have developed through several phases: browsing static web pages, using mutual web tables, processing B2C commerce, till developing B2B application integration. More than half of the enterprises still rest on the phrases of static web pages and simple e-commerce. Up to present, the main form of e-commerce is B2C, which exhibiting its information through HTML, and many of its processes is launched by manual work. With the emergence of B2B and other more complicated commerce pattern, tradition e-commerce technique pattern could not guarantee for validity and could not complete tasks in time. E-commerce of next generation demands for automatic Web dealing initiated by application, decrease of error easily arisen in manual process, and efficiency. Automatization, intelligentization and integration should be embodied by the e-commerce technique pattern in next generation. More research in collaborative information technique and distributive computing pattern in new generation is the main approach to solve the problem. It will promote the development of collaborative e-commerce system that is the third generation of e-commerce. Due to its connatural advantage, mobile agent is applied in next generation e-commerce as a new technology.

Therefore, the primary aim of this paper is to give an architectural model and a framework for implementing mobile agent-based e-commerce system. Intelligent and mobile agents are applied to current e-commerce technology to meet autonomy, intelligence and integrity requirements. For the special requirement of electronic commerce, system security is the one of the most important elements to a successful system. The original idea is to establish a secure mobile agent system (SMAS) firstly, and then to build up the mobile agent-based e-commerce system framework based on the SMAS to ensure transmission security.

The remainder of this paper has the following structure: Section 2 gives an introduction to mobile agent firstly, and then establishes a secure mobile agent system

model for e-commerce. In the last portion of this section, a mobile agent-based e-commerce system framework is brought forth. A logic verification of the system framework is demonstrated in section 3. Section 4 describes some issues relevant to mobile agent applied in e-commerce. And the section 5 gives a conclusion.

2. MOBILE AGENT-BASED E-COMMERCE

2.1 Mobile Agent Introduction

The Object Management Group defines a software agent as "a computer program that acts autonomously on behalf of a person or organization"^[1]. The following properties characterize agents:

- pro-active (support of the user's work)
- adaptive (learning the user's preferences or the ability to work on different platforms)
- autonomous (limited communication with its creator)
- intelligent (making 'intelligent' decisions)
- mobile (can actively migrate in networks to different systems and move directly to the local resources, like databases or application servers)

Mobile agents are programs that can migrate from host to host in a network, at times and to places of their own choosing. The state of the running program is saved, transported to the new host, and restored, allowing the program to continue where it left off. Lange and Oshima^[2] give reasons why to use agents: e.g. improvements in latency and bandwidth of client-server applications and reducing vulnerability to network disconnection.

2.2 Mobile Agents in E-Commerce

Mobile agents are well suited for electronic commerce. A commercial transaction may require real-time access to remote resources such as stock quotes and perhaps even agent-to-agent negotiation. Different agents will have different goals, and will implement and exercise different strategies to accomplish these goals. We envision agents that embody the intentions of their

creators, and act and negotiate on their behalf. Mobile agent technology is a very appealing solution to this kind of problem.

An electronic commerce transaction may be viewed in terms of four different phases [3]: product brokering, merchant brokering, negotiation, payment and delivery.

Product brokering consists in the gathering of information about the product that is going to be bought.

Merchant brokering involves the evaluation of a set of alternatives in order to make the purchase. Making the decision implies considering all the tradeoffs that the various products offer: price, warranties, delivery time, and others.

During the negotiation phase the agent settles the final terms of the commercial transaction. The characteristics of the market directly influence the outline of this phase. In markets where prices and characteristics are fixed, negotiation may not even exist.

Finally, in the purchase and delivery phase of the transaction, the agent actually makes the acquisition and delivers the money (or its electronic equivalent) against the goods.

2.2.1 Secure Mobile Agent System

A number of advantages of using mobile code and mobile agent computing paradigms have been proposed [4, 5, 6]. One of the main obstacles to the widespread adoption of mobile agents is the legitimate security concern of system developers, network administrators, and information officers. It has been argued that once the security issues have been resolved and a collection of security mechanisms have been developed to counter the associated risks, then the users of mobile agent technology will be free to develop useful and innovative solutions to existing problems and find a wide array of application areas that will benefit from this technology. Using this collection of security mechanisms to mitigate agent-to-agent, agent-to-platform, and platform-to-agent security risks may, however, introduce performance constraints that could dictate design decisions or negate the benefit of using mobile agents for certain applications. Considering the special security requirements in e-commerce, a secure mobile agent system (SMAS) framework is presented in this paper.

SMAS consists of mobile agents and agent server (e.g. agent facilitator). Mobile agent has six modules: security shell, environment interaction module, state denotation module, task execution module, routing policy module and log file. The architecture of mobile agent is shown in figure 1.

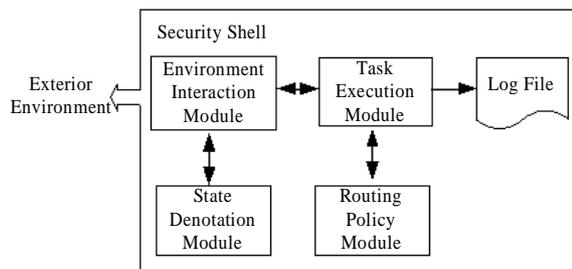


Figure 1 Mobile Agent Architecture Model

Security shell is responsible for protecting the agent from attacks of malicious hosts and other agents. The main functions include encrypting and decrypting the whole agent, authorizing the visit host, deploying the different security policy according to the different trusted level of host.

Environment interaction module is responsible for the maintenance of communication between agent and exterior environment that includes sensor (apperceive outside environment), message processor (process inter-messages, data communication between agent and environment based on KQML) and reactor (i.e. effector, output the result to outside environment).

Task execution module is the execution module of agent, which consists of action execution module and result integration module. According to the security strategy to different methods and data deployed by security shell, task execution module executes corresponding method, and makes integrated analysis of results for its goal.

State denotation module is responsible for recording the state of agent, which includes attribute value of agent and the sequence of agent's running environment. This makes agent to run successfully when it is hold up or resumed. Agent's state can also be recorded in fixed interval for its resuming in abnormal circumstance.

Routing policy module plans agent's migration route. There are two feasible routing policies, one is fixed routing, and the other is dynamic routing based on formula and catalog services.

Log file records every sensitive instruction and manipulation on agent executed by agent platform, it's used for audit afterwards. Log file must be encrypted so that it couldn't be juggled.

Mobile agent server (mobile agent service facilitator) is based on agent transfer protocol (ATP), it provides essential EE(execution environment) for agent's migration, execution, and other functions and services such as dispatch, reception, recovery, security management and service transfer. Mobile agent server includes security management, agent executing environment, agent API, agent service environment, communicating API, as shown in figure 2.

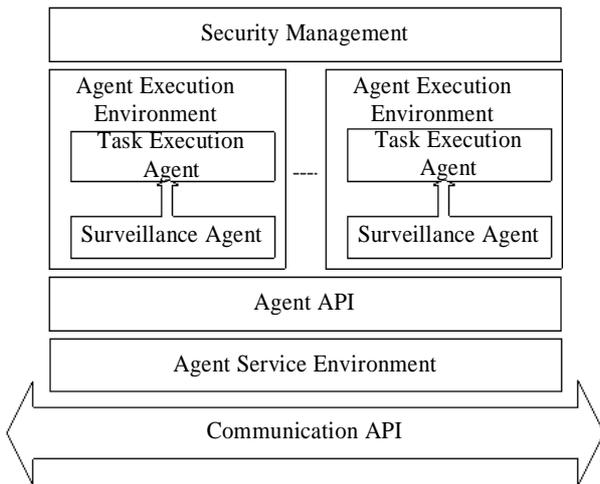


Figure 2 Mobile Agent Server Structure

Security management is responsible for the security of agent server and the agent executing on it, it takes authentication, encryption, visit controlling and other measures to protect system security.

Every agent server can process several agent EEs. Agent EE is the resource distributed for every agent by the agent server. It runs agent under the control of security management module. Every EE has two kinds of agents: task execution agent and surveillance agent. Task execution agent runs the program on behalf of agent user. Surveillance agent keeps an eye on the execution situation of the task agent. It can terminate the task agent once there is any unauthorized or illegal operation.

Agent API is the unique interface between agent and agent server. It provides many characteristic operations of mobile agent that can be utilized to finish the distributed task for the agent. For instance, agent calls Jump procedure explicitly before migrating to another host. Jump procedure captures the whole state image of the mobile agent and transfers the agent to the destination host. On the destination host, agent server loads the agent to the relevant agent EE and recovers the agent from the hold point. The primary agent will be terminated on the local host after the agent's state image being captured, transferred and recovered successfully on the destination host.

Agent service environment manages lifecycle, event and persistence for agent. Lifecycle management realizes creation, dispatch, migration, reception, storage and termination for mobile agent, event service provides agent communication mechanism for collaboration with other agents or application system, which supplies essential condition for agent collaboration. Persistence service ensures agent execution's persistence through corresponding mechanism, it helps agent to restart accurately when there is system or network disaster before agent reach its destination.

Communication API is responsible for transmitting mobile agents, data and messages, including network protocol API and email API.

2.2.2 Mobile Agent-based E-Commerce

Based on above construction of mobile agent system, this paper brings forth the mobile agent-based e-commerce system framework. Its model is shown in figure 3.

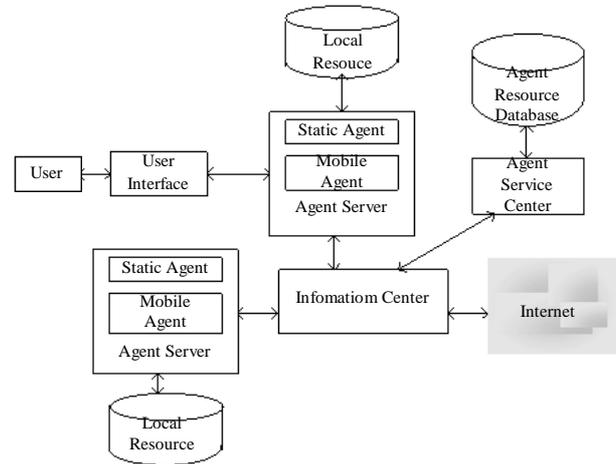


Figure 3 Mobile Agent-Based E-Commerce System Framework

In this system, agent service center provides catalog service, registers mobile agent's basic information, maintains agent's lifecycle, cancels illegal agent in time and provides agent query service. In order to enforce system security, especially for mobile agent protection, agent service center also maintains a table which records the history of host visitation, it gets task accomplish information of agent in visitation from its log files, calculates the host's credit degree, puts the result in the history table and re-calculates it so that mobile agent could optimizes routing strategy. Service center need to provide host service query function in order to help agent affirm the servers on which its task could be accomplished, and choose the best route.

Information center is responsible for security control of the whole E-Commerce system, it generates certain inbreak inspection agents, sends them to each sub-system, monitors real time condition of system execution, collects relative information of inbreak event, analyzes inbreak event, traces inbreak route and confirms its beginning. Control center takes corresponding measure when it received inbreak report and analyzed it. Control center is also responsible for downloading patches of application and operation system, upgrading virus-preventing software, after this it dispatches mobile agent to install them on every node in the system.

Following is the workflow of system:

- (1) User submits a request through user interface.
- (2) Agent Server analyzes the request and creates static agent or mobile agent according to the request of task, then registers the agent's information in agent service

center.

(3) Static agent visits local resource under security control of agent server, completes user's request and returns the result to user. In the meantime, it registers agent's logout information on agent service center.

(4) Mobile agent registers its information on agent service center, queries for host service table, returns agent server, works out migration route or plan according to routing strategy.

(5) Mobile agent queries the history table of host visiting on agent service center, gets credit degree of host on the route, returns the optimized migration route of agent server.

(6) Mobile agent sends out migration request to server, agent server makes agent hung up, encrypted, packed and sends it to destination host.

(7) Destination host accepts agent, authenticates each other.

(8) Destination host carries out security strategy according to authentication result, assigns limited execution environment for agent, and generates monitor agent to monitor the execution information of mobile agent.

(9) Mobile agent carries out different security strategy according to authentication result, executes task under assigned resource. If necessary, it could make alliance with other agents and accomplish the task collaboratively.

(10) Agent goes through hosts in routing table and completes its task, collects the results, returns it to user and submits log files to agent service center, registers logout information.

3. VERIFICATION

At present, many technology research and system development based on agent have been gotten along at home and abroad. However there is little research on agent system verification. Yanxiang He and Xinmeng Chen introduced an agent-mediated system verification method in 2001^[11]. The definition of agent-mediated system verification can be described as follow: a system contains one or more claimed agent(s), it will be changed when some other claimed agent(s) add(s) to the system. That means:

(a) System behavior transformation doesn't indicate other process's running slows down the system processing speed. Otherwise it shows the claimed agents apperceive each other to a degree, and they adjust their actions accordingly.

(b) New added agent must have the same style with the claimed agents, viz., although it has different goal, knowledge and belief, it should have the same architecture and function.

(c) The existing agents and the new added agent don't always communicate with each other, and they not always be autonomous, persistent or intelligent.

According to the definition, following test should be done. We suppose X have two instances. If the behavior

when there is another instance to exist of each instance and its behavior while there are no other instances to exist are different, that indicates X is an "agent". On the contrary X is not an "agent". The verification has nothing to do with agent's mobility.

For example, a user submits a purchase requirement, two instances of information collection agent, agent1 and agent2, are created. Agent1 and agent2 move in the network according to their migration route and collect the required commodity information. Each agent apperceive environment automatically and make action according to the environment. When one agent's reaction reaches the existence of another agent, they communicate with each other and get the new task result and execution decision by negotiation. Then they regulate their behaviors based on the new strategy. So they are "agent" on the agent-mediated system verification definition mentioned above. The rest may be deduced by analogy, we can prove that the system established on the framework introduced in section 2 is an agent-based system.

This verification method is based on agent's sociability. For security and performance verification of the system, demonstration and simulation verification are needed.

4. RELATED AND FUTURE WORK

Security insurance and system reliability are the foundation and key for the popular application of the mobile agent-based e-commerce. In this paper, secure mobile agent architecture is designed for protecting agent from attacks of malicious host and other agents; surveillance agents are added for protecting the host from attacks of malicious agents; XML-PKI technology is provided for encrypting the sensitive data, code and states for transaction security of agent. Disaster recovery technology is applied to buildup system reliability.

There are many promising new research on mobile agent technology and its applications in recent years^[8,9,10]. However we should see that we must solve some more technology obstacles on our way. The first one is the issue of agent standardization. FIPA and MASIF are working on that and there should be some standards about mobile agent system framework and so on. For agent in e-commerce, it is still in its infancy. The main trend in this area is to attack four problems: negotiation, information discovery, ontology and security. Furthermore, the use of electronic money in conjunction with mobile agent is particularly challenging. The integration of digital payment system into mobile agent-based e-commerce is important to realize the potential e-market. In a word, there are many jobs should be done for the mobile agent-based e-commerce application in the future.

5. CONCLUSION

Mobile agent-based e-commerce system framework mentioned above utilizes the virtues of mobile agent technology fully. It can realize the electronic transaction automatically and intelligently. It also can provide the better security protection for effective and convenience business cooperation and collaborative information sharing between enterprises. To realize the system concretely, Java language is adopted to implement the mobile agent for platform independence. Mobile agent communicates in KQML (Knowledge Query and Manipulation Language) with each other and accesses services provided by agent server. Mobile agent server adopts ATP (Agent Transfer Protocol) to dispatch and receive mobile agents.

ACKNOWLEDGEMENT

The research work has got a lot of people's help and energetical support. Especially thank professor Yang for his guidance and support.

REFERENCES

- [1] Crystaliz, Inc., General Magic, Inc., GMD Focus, and IBM Corp, "Mobile Agent Facility Specification Technical report", *OMG*, 1997.
- [2] Danny B. Lange, Mitsuru Oshima, "Seven Good Reasons for Mobile Agents", *Communications of the ACM*, Vol 42, No. 3, pp88-89, March 1999.
- [3] P. Maes, R. Guttman, and A. Moukas, "Agents that Buy and Sell", *Communications of the ACM*, vol. 42, 1999.
- [4] Danny Lange, Mitsuru Oshima, "Programming and Deploying Java Mobile Agents with Aglets", *Addison-Wesley*, 1998.
- [5] Jonathan Smith, "A Survey of Process Migration Mechanisms," *Operating Systems Review, ACM Special Interest Group on Operating Systems*, Vol 22, No. 3, pp28-40, July 1988.
- [6] David Chess, Colin Harrison, and Aaron Kershenbaum, "Mobile Agents: Are They a Good Idea?", *IBM Research Report RC 19887 (88465)*, December 1994.
- [7] Christian F, Tschudin, "Open resource allocation for mobile code", *In Proceedings of The First Workshop on Mobile Agents*, pp186-197, April 1997.
- [8] Birk A, "Boosting cooperation by evolving trust", *Applied Artificial Intelligence*, Vol 14, No. 8, pp769-784, 2000.
- [9] Castelfranchi C, Falcone R, "Trust and control: dialectic link", *Applied Artificial Intelligence*, Vol 14, No. 8, pp799-823, 2000.
- [10] Zacharia G, Moukas A, and Maes P., "Collaborative reputation mechanisms for electronic marketplaces", *Decision Support Systems*, Vol 29, pp371-388, 2000.
- [11] Yanxiang He, Xinmeng Chen, "Design and application of agent and multi-agent system", *Wuhan University Press*, pp203-235, 2001.

Towards Collaborative Travel Recommender Systems

Cane Wingki Leung, Stephen Chifai Chan, Korris Fulai Chung

Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
{cswkleung, csschan, cskchung}@comp.polyu.edu.hk

ABSTRACT

Collaborative filtering (CF) based recommender systems have been proven to be a promising solution to the problem of information overload. Such systems provide personalized recommendations to users based on their previously expressed preferences and that of other similar users. In the past decade, they have been successfully applied in various domains, such as the recommendation of books and movies, where items are simple, independent and single units. When applied in the tourism domain, however, CF falls short due to the simplicity of existing techniques and complexity of tourism products. In view of this, a study was carried out to review the research problems and opportunities. This paper details the results of the study, which includes a review on the recent developments in CF as well as recommender systems in tourism, and suggests future research directions for personalized recommendation of tourist destinations and products.

Keywords: Recommender System, Collaborative Filtering, Personalization, Data Mining, Travel & Tourism

1. INTRODUCTION

In the past decades, the rapid development of Internet technologies has led to severe information overload. The amount of information available on the web, and the amount added daily, is too much for us to process. The use of recommender systems has therefore become increasingly popular. Such systems provide advices to users about items they might like to purchase or examine, avoid them from being drown in the vast amount of information [3].

There are three major types of recommender systems: content-based, knowledge-based and social- or collaborative-filtering (CF) based. Content-based recommender systems establish users' interest profiles by analyzing the features of their preferred items. In such systems, features of items are compared to those of the preferred items of the user. The more relevant items, measured by features similarity, are recommended. Knowledge-based recommender systems make use of knowledge about users and products to generate recommendations. They use a reasoning process to determine what products meet a user's requirements. Social-filtering or CF based recommender systems provide personalized recommendations according to user preferences. Such systems maintain data about users' purchasing habits or interests and from there identify groups of similar users. For a target user, known as the active user, items liked by other similar users are recommended.

In addition to providing personalized recommendations, CF offers a number of advantages over the other two techniques, making it one of the most promising solutions to information overload. As no content information is considered in the recommendation process, items being filtered need not be amenable to parsing by a computer. Besides, recommendations generated are based on tastes of users rather than more objective properties of the items themselves. The items

recommended to a user can therefore be very different (content-wise) from what the user has liked previously, overcoming a major limitation of content-based recommender systems [31]. When compared to knowledge-based recommender systems, CF techniques are much simpler and easier to implement. While the former requires a domain knowledge engineering process to build a knowledge base of items, the later can be fully automated. On the one hand, this allows CF to be easily applicable to any domain where a database of user preferences is available. On the other hand, this makes CF falls short when recommending complex items such as travel purchases [4]. Nevertheless, CF has become the most successful technology applied in recommender systems. To date, it has already been studied and applied to various domains such as the recommendation of books, Usenet articles and jokes.

In recent years, CF has gained research attention in the tourism domain for several reasons. By observing and predicting user preferences, CF prevents users from being overwhelmed by vast amount of travel information. This is particularly valuable in an information-intensive domain like tourism, which comprises a large variety of products such as accommodations, attractions, restaurants, and many others. Besides, one characteristic of the purchase of tourism products is that at the moment of decision-making, only information about the product, not the product itself, is available to users [26][33]. Current web-based tourist information portals can present archives of pre-defined destinations, some with rich multimedia content. While tourism product suppliers are always in favor of their own businesses, other visitors' experiences would probably provide more valuable and unbiased information about a destination. CF techniques would help users obtain such information according to their travel preferences [4].

Yet, CF is not readily applicable to the recommendation of tourism products [4][28], which are more diversified

and complex than products like books. With the objective to bridge the gap between CF and tourism products recommendation, a study was carried out to identify the problems in recommending tourism products, and to review the state-of-the-art in both areas. This paper details the results of the study, and suggests future research directions to work towards collaborative travel recommender systems.

1.1 Collaborative Filtering (CF)

The underlying philosophy of CF is that each individual belongs to a larger group of like-minded individuals. CD-based systems therefore make personalized recommendations to users based on the interests of similar users [7][20][27][31]. In a movie recommender system, for example, the movie Shrek may be suggested to a user Sam, after noticing that his taste is similar to that of another user Bob, who has liked Shrek.

Users' preferences provide the basis for making collaborative recommendations. Such preferences can be expressed explicitly by means of ratings, or implicitly such as in purchase history. CF algorithms may compute a predicted score (rating) of a particular item or a list of recommended items, known as Top-N recommendation, for the active user.

1.2 Recommending Tourism Products

Recommending tourism products poses additional challenges to existing CF researches. They are mainly caused by the complexity of tourism products, and the special taxonomies they can have.

Firstly, tourism products are highly heterogeneous even at a single destination. They include restaurants, hotels, theme parks, and many others. Recommending these products with pure CF techniques can be easy, as long as each item is uniquely identified (by a product code, for example) and no content information is considered. However, these make less sense for tourism products as content information may also be essential in the filtering process. For example, if a user is looking for a place to ski, a travel recommender system must be able to filter out irrelevant activities such as hiking. Some existing techniques such as [14] and [30] already allow content information to be used in the recommendation process. This, however, can be more difficult in recommending tourism products due to **product heterogeneity** since products can have different representations, and the filtering of content information may be based on different features for different products.

Furthermore, tourism products may also contain essential context information such as availability and location. For example, if a user is looking for a place to ski in June, a travel advisory system may consider Australia as a suitable destination. If his/she wishes to ski in December, however, it will be more suitable to recommend some places in Europe due to the availability of the desired activity at different locations

and points in time. **Context awareness** is therefore another important factor in designing a travel recommender system.

Lastly, tourism products have strong inter-item relationships. They may also be arranged into different hierarchies or taxonomies. For example, the is-a hierarchy may exist if products are categorized according to their type (e.g. Surfer's Paradise is a beach). Besides, the part-of or component hierarchy may also exist if they are categorized according to their location (e.g. Surfer's paradise is part of Gold Coast). Another special example of component hierarchy in tourism products is tour package, which comprises multiple heterogeneous products such as flight tickets, accommodation, as well as tourist spots (e.g. a visit to Surfer's Paradise is included in a 2-day tour to Gold Coast). These products may exist in multiple packages, or may be purchased alone as a single unit of product. To recommend tourism products, different **product taxonomies** should be taken into consideration.

1.3 Organization

The rest of this paper is organized as follows. In Section 2, different types of CF techniques are described and their strengths and weaknesses are commented on. In Section 3, several CF applications are introduced. Related work in tourism is also described. In Section 4, some CF techniques and related work introduced are further discussed with respect to their ability to recommend tourism products. From there some future research directions are identified. Lastly, Section 5 concludes this paper.

2. CF TECHNIQUES

CF algorithms were traditionally classified into two major types, namely memory-based and model-based [2]. Memory-based algorithms operate over the entire user database, and recommend products using statistical methods. Model-based algorithms construct compact models from the user database, and recommend products using probabilistic methods. Since such models can be constructed offline, model-based algorithms usually improve online performance.

Such classification mainly reflects how products are recommended in real-time – either based on the entire database or models constructed and stored in memory. To more specifically describe various CF algorithms, in this paper they are classified by the underlying techniques they used to make recommendations.

2.1 K-Nearest Neighbor (k-nn)

K-nearest neighbor (k-nn), also known as pure CF, was commonly used in early CF-based systems. It works in three major steps, namely similarity weighting of users, neighbor selection and prediction computation.

At the similarity weighting stage, *each user* in the

database is assigned a similarity weight using some measures (such as those analyzed in [2] and [27]). Such similarity is reflected in the ratings they have given on items. For two users to be comparable, only items that both users have rated are counted. At the neighbor selection stage, a number of k nearest neighbors, who are the users having the highest similarity weights, of the active user are selected as predictors for items. Based on the interests of the selected neighbors and some partial information of the active user, prediction score of an item is computed.

K-nn makes predictions that are ratings-based and content-independent. It can therefore be applied to domains where textual descriptions of products are not available, not meaningful, or cannot be easily categorized by any attribute. Furthermore, it uses statistical approaches that are easy to implement, and can be applied as long as a database of users preferences on items is available. Despite its simplicity, empirical analyses [2][12] proved that it produces good predictions, especially if the active user has rated a significant number of items. Besides, predictions are generated by performing real-time computations over the database. As no pre-compilation of data is required, new data points can be easily added to the application.

For the above reasons, k-nn was commonly used in recommender systems. This, however, has revealed a number of challenges for researchers. Among which the most essential ones are data sparseness, cold start problem and scalability.

Data sparseness arises as users can only rate the items, or part of the items, that they have access to (that is, they cannot rate any unobserved items). In k-nn, it is possible that the set of common items rated by the active user and his neighbors is very small, even if the selected neighbors are already the "nearest" ones. This resulted in poor prediction quality [29].

Furthermore, data sparseness implies that given a very large database of items, there may not be a complete set of ratings across all items. The cold-start problem, or the first-rater problem, arises when no prediction about a certain item can be made for the active user as no other users have rated it before [16].

Scalability is an important issue especially in real-time applications. K-nn has severe performance bottleneck in the similarity weighting stage, as it requires real-time computations to be performed over the entire database for each prediction. Performance of the algorithms degrades when the numbers of users and items grow in the system.

2.2 Classification

Some researches regarded CF as a classification problem [1][21]. Instead of predicting the rating of an item, classification techniques predict whether the active user would like or dislike it. To do so, user ratings are

first discretized, usually into two classes - *Like* and *Dislike*. Classification techniques, such as Bayesian network [2], are then used to make predictions for the active user.

A Bayesian network makes use of decision trees to predict the likeliness of items, conditioned on whether some other items have been liked previously. Since decision tree induction is beyond the scope of this paper, it is not described here. Please refer to [11] for details.

In this technique, decision trees induction is done offline. Only the decision tree built for the target item has to be inspected in real-time, resulted in improved online performance over k-nn. Prediction quality, however, degraded due to its probabilistic nature.

2.3 Clustering

2.3.1 User-based Clustering

The Bayesian cluster model divides the user base into many groups (clusters) according to their similarity [2]. The recommendation process first assigns the active user to the cluster containing the most similar users. Then, it uses the preferences of users in that cluster to generate recommendations. Since the active user is only compared to a controlled number of clusters, scalability is improved. Similar to the classification technique, it produces less personal recommendations.

Eigentaste [8], another user-based clustering technique, clusters users by their similarities reflected in the ratings they have given on a gauge set of items. Then, the mean for each non-gauge item in each cluster is computed based on the number of users who have rated that item. Lookup tables of recommendations for each cluster are built by sorting the non-gauge items in descending order of their mean ratings. In real-time, recommendations are obtained by querying the pre-computed lookup tables. This simple online process ensures scalability. Since users are required to rate a gauge set of items, the ratings matrixes representing their preferences are dense. Users can therefore be well correlated to preserve prediction accuracy.

2.3.2 Item-based Clustering

Sarwar et al. [29] explored the item-based clustering approach, in which recommendations are produced by finding items that are similar to those the active user has liked previously. As the number of items and their relationships are relatively static, as compared to the number of users and ratings, recommendations can be computed offline. This avoids the performance bottleneck of k-nn and improves scalability. However, prediction quality of this approach only shows slight improvement over k-nn. This might be due to the fact that data sparseness was not addressed in the research.

2.4 Association Rule Mining

Association rule mining (ARM) is a data mining technique for discovering interesting relationships between items by finding items frequently appeared together in transactional databases. An association rule is denoted as $A \Rightarrow B$, and measured by two measures known as support and confidence. For example, the rule $A \Rightarrow B$ [20%, 90%] means that 90% of users who purchased item A also purchased item B, and 20% of all users purchased both of them [11].

ARM is usually used for market basket analysis. It can also be used to generate recommendations after interesting rules are mined. For example if $A \Rightarrow B$ is interesting in a certain system, and the active user has liked item A previously, he/she will probably like item B also. Note that a rule is considered as interesting only if its support and confidence values are higher than the user-specified minimum.

2.4.1 Adaptive Association Rule Mining

Lin et al. [18], however, pointed out that traditional ARM algorithms are inefficient for collaborative recommendation for two reasons. Firstly, such algorithms mine rules for all items in the database. Therefore, many rules mined will not be relevant for a given user. Secondly, the minimum support and confidence have to be specified in advance, which might lead to either too many or too few rules due to variations in user tastes and items' popularities.

To address these issues, they suggested an adaptive-support rule mining algorithm, known as AR-CRS, tailored for CF. AR-CRS mines rules for one target item at a time, and automatically adjusts the minimum support value to mine a user-specified number of rules. It can compute associations among users, items, or both.

2.4.2 Multi-level Association Rule Mining

Kim & Kim [15] applied multi-level association rules mining (MAR) to address data sparseness and the cold start problem. MAR can be applied on items that are organized in a hierarchical category structure, such as the classification of goods in a department store (is-a hierarchy). Its key feature is to mine rules at different level of abstraction [10]. A major concern of MAR is that higher support is likely to exist at higher level of abstraction. Different minimum supports should therefore be used at different levels.

When applied in CF, MAR can compute preferences for items that are not covered by the traditional, single-level technique. To do so, the category C_i to which a preferred item i belongs is found. If any association rule among C_i and another category C_j exists, some preferences will be given to items that belong to C_j . The consideration of such product taxonomy allows recommending cold-start or less popular products.

2.5 Hybrid Techniques

Personality Diagnosis (PD) [25] combined k-nn and clustering to improve prediction quality. In addition to the traditional similarity weighting measures, it employs a probabilistic interpretation of results to provide better predictions. PD retains the advantages of k-nn such as simplicity and adding new data points with ease. In terms of accuracy, it outperforms several approaches including k-nn and the Bayesian network model although it does not address data sparseness. In terms of time and space performance, it does not provide any advantage over k-nn.

O'Conner & Herlocker [23] combined k-nn and clustering to improve scalability. Their technique uses existing data partitioning and clustering techniques to cluster items based on user ratings. Predictions are then computed independently in each cluster. Experimental results show that their technique is more scalable than k-nn, as its real-time prediction process only considers items in a single cluster. Consistent improvement, however, in its prediction accuracy could not be observed. A possible reason for this is that their algorithm restricts items to being exclusively in one single cluster. Items that might have significant predictive value for multiple clusters therefore could not reflect real user preferences.

Sarwar et al. [30] combined information filtering (IF) and CF to provide better recommendations. They integrated filtering agents, known as filterbots, that evaluate and rate items using syntactic features to provide a dense ratings set. This improves prediction accuracy, and its effectiveness was proved in [9].

The Item-based Clustering Hybrid Method (ICHM) also integrated content information into CF to improve prediction quality and solve the cold start problem [17]. In ICHM, items are clustered based on content information. In a movie recommender, for example, movies are clustered by their genre, director, actor, etc. For a new item with no ratings, its prediction score is computed as a weighted sum of the ratings given to other items in the same cluster. Experimental results show that this approach improves prediction accuracy. It is also able to recommend cold-start items with reduced accuracy, as compared to using a dataset with no cold-start items.

3. CF APPLICATIONS AND RELATED WORK

CF has been proven successful in both research and practice. A few examples selected from different domains are introduced in this section. Amazon.com [19] is one of the most successful and well-known online retailers. It makes use of purchase histories of customers to produce recommendations using the item-based technique. GroupLens [27][16] is a pioneering and well-known project in automated CF. It had used k-nn to recommend Usenet articles. Filtering agents were later integrated into the system to improve prediction quality [30]. MovieLens is another recommender system developed by the GroupLens team. It recommends

movies to users using GroupLens' technologies. Jester [8] is an online joke recommender. When it was first developed, it had used k-nn to generate recommendations. When the number of registered users increased, however, it was found that the processing time of the system increased linearly until it crashed. The constant time algorithm Eigentaste was therefore developed and adopted.

Some related applications in the tourism domain were also studied. They are mainly based on knowledge-based technologies. CF is only used to rank the recommended products according to user preferences.

Dietorecs [4] and NutKing [22] are two similar applications developed by the eCommerce and Tourism Research Laboratory (eCTRL) [6]. They recommend travel products and activities based on the Case-Based Reasoning (CBR) technology, which is a problem solving technique that solves the current problem by adapting solutions for previously solved, similar problems. The CBR cycle consists of 4 stages, namely Retrieve, Reuse, Revise and Retain. Detailed descriptions of each stage can be found in [34].

A case base refers to a repository of representative cases, which are previously adopted solutions and/or possible solutions to new problems. In Dietorecs and NutKing, it is actually a pool of travel products and activities recommended by domain experts with respect to different travel settings (e.g. duration of travel). To start generating recommendations, a new case containing the active user's profile and travel settings is created. Pre-defined recommendations are then retrieved from the case base (the Retrieve stage), and ranked according to the preferences of the active user. The retrieved cases can be accepted as-is, or modified by the active user by adding or removing elements in the recommended cases (the Reuse stage).

Ski Europe [32] recommends places for skiing in Europe based on the TripMatcher technology [4]. Similar to Dietorecs and NutKing, the system maintains a knowledge base of destinations, developed by professional researchers worldwide. For each destination, travel experts have ranked different activities during different times of the year. TripMatcher applies information filtering and text mining algorithms to such context information, products' content information and users' ratings to generate recommendations.

In CBR, management of the case base is an important issue as it directly affects the correctness and quality of the solutions produced by a system. The major concern here is the selection of representative cases, including the number of cases and their content. In the CBR cycle, the Revise and Retain stages allow domain experts to, based on the solutions modified by users in the Reuse stage, revise the cases originally stored in the case base, retain the cases by updating the original ones, or store them as new cases. Thus they are very important steps

to help maintain the quality of the case base. However, how these important steps have been addressed in Dietorecs and NutKing, is unclear. Although the TripMatcher technology does not rely on CBR, it faces a similar problem as it produce recommendations from a knowledge based of pre-defined trips.

Due to the complexity of tourism products, it is important to develop a data model for their representation. As mentioned before, tourism products are heterogeneous, complex, inter-related and may contain essential context and content information. However, existing CF algorithms are all based on very simple data models. In fact, only ratings/purchase data and the unique identifiers of products are used in most algorithms. Integrating data models in CF would facilitate the development of more sophisticated, flexible and interoperable recommendation engines.

The Open Tourism Consortium (OTC) [24] is a consortium of companies, government agencies, individuals, and universities participating in the development of publicly available standards and open source software to support tourism. Recently, they developed an XML-based data exchange language known as TourML to represent objects and events, such as hotels and restaurants, of interest to tourists. Although OTC is not doing any CF-related research, their work can provide reference for modeling different tourism products.

4. FUTURE RESEARCH DIRECTIONS

As aforementioned, the widespread use of pure CF methods has revealed some fundamental challenges concerning the scalability and prediction quality of CF systems. Since then, CF researches have been focusing on developing more scalable and accurate algorithms. Although numerous approaches have been developed, the dominant paradigm for performing CF has been based on techniques regarding items in the domain as some simple, single and independent units. This paradigm, although has been proven successful in various domains, falls short when applied in the tourism domain which poses additional challenges namely product heterogeneity, context awareness and the existence of product taxonomies to CF.

Another problem in this paradigm is non-transitive association [14]. In user-based methods, if two users have both experienced or rated similar items but not the same ones, their correlation is lost. This is known as the user-based non-transitive association. In contrast, using item-similarity instead of user similarity avoids this problem, but results in the item-based non-transitive association problem if two similar items have never been experienced or rated by a user.

Due to the presence of content and context information, purely CF algorithms may not be sufficient for travel recommender systems. Existing solutions, such as TripMatcher [4], therefore combined different

techniques with CF to form a recommendation engine. Such technology has already addressed product heterogeneity and context awareness using information retrieval and text mining algorithms. However, it does not take into consideration product taxonomies in the recommendation process.

In fact in most CF algorithms, recommendations are generated based on the relationship between some attributes of items and users' profiles. The item-based technique [29] behaves differently in the technical aspect, but still it is based on the relationship between items and users identified earlier to generate recommendations. They do not take into consideration inter-item relationships and product taxonomies, which are likely to be found in tourism products. Obviously, treating items as single units in a recommender system is inappropriate in the tourism domain.

MAR has already shown that it is possible to integrate the is-a hierarchy to improve recommendations [15]. Such hierarchy, and other possible hierarchies, have not yet been considered in travel recommender systems. This reveals a potential research problem, which is to extend the existing techniques to incorporate different product taxonomies in the recommendation process. This may lessen the problems of non-transitive association in CF, because item similarities are already implied in the hierarchy itself.

Yet, MAR applied to the is-a hierarchy may not be readily extendable to other hierarchies such as part-of. The is-a hierarchy exercises the upward closure property which affects the generation and pruning of associations rules. In such hierarchy, the support count of a higher-level item must be equal to the sum of that of its child items. In other words, if a higher-level item is infrequent (with respect to the user-specified minimum support), all of item child items must also be infrequent. In contrast, if a lower-level item is frequent, all of its parent items must also be frequent. Due to the presence of this property, the algorithm in [15] recommends products of a certain category, if some associations between that category and the category of a previously liked item exist. Considering other taxonomies in the recommendation process, however, have not yet been studied in the literature of CF. Further research on this is required.

To apply MAR in travel recommender systems, its challenges and shortcomings should also be addressed. The major concern in MAR has been the determination of minimum supports at different level of hierarchies (is-a hierarchy). Whether this applies to other product hierarchies is unknown, but should be aware of in future researches. The adaptive minimum support method proposed in [18] may be considered, as it can adjust the minimum support value according to the desired number of resulting rules.

A shortcoming of applying ARM/MAR in CF was observed. Traditional ARM/MAR generate "interesting

rules" based on items' frequencies of appearance in transactions in a database. In other words, the appearance of an item in a user's transaction implies that it is liked by that user. In classical ratings-based CF, however, ratings can be positive or negative. Using ARM/MAR requires transforming ratings data into class labels such as like and dislike. In [15], for example, a certain rating given by a user is converted to 1 if it is greater than the average rating given by that user, or 0 otherwise. Applying this technique to the Jester dataset [13], ratings should first be transformed from a continuous rating scale (-10 to 10) to like (1) and dislike (0). Consider a case where all ratings given by a user are negative with an average -5. According to [15], a joke with rating -1 will also be considered as being liked by the user. This actually implies a loss of information in the user's likeliness of items. Fuzzy logic may be a possible solution to this problem.

5. CONCLUSION

CF has been a promising solution to the problem of information overload in the past decade. In recent years, it began to gain research attention in the tourism domain for its ability to produce personalized recommendations. However, existing CF techniques fall short when recommending tourism products which are heterogeneous and complex. With the objective to bridge the gap between CF and tourism products recommendation, a study was carried out to identify the problems in recommending tourism products, and to review the state-of-the-art in both areas. This paper details the results of the study, and suggests future research directions to work towards collaborative travel recommender systems.

REFERENCES

- [1] Basu, C., Hirsh, H., Cohen, W., "Recommendation as classification: using social and content-based information in recommendation". *Proceeding of the 15th National Conference on Artificial Intelligence*, pp. 714-720, 1998
- [2] Breese, J. S., Heckerman, D., Kadie, C., "Empirical analysis of predictive algorithms for collaborative filtering". *Proceeding of the 14th Conference on Uncertainty in Artificial Intelligence*, pp. 43-52, 1998
- [3] Burke, R., "Knowledge-based recommender systems". In A. Kent (ed.), *Encyclopedia of Library and Information Systems*. Vol. 69, Supplement 32. New York: Marcel Dekker, 2000
- [4] Delgado, J., Davidson, R., "Knowledge bases and user profiling in travel and hospitality recommender systems". *Proceeding of ENTER 2002*, pp. 1-16, 2002
- [5] Fesenmaier, D. R., Ricci, F., Schaumlechner, E., Wöber, K., Zanellai, C., "Dietores: travel advisory for multiple decision styles". *Proceeding of ENTER 2003*, pp. 232-241, 2003
- [6] eCommerce and Tourism Research Laboratory: <http://ectrl.itc.it>
- [7] Goldberg, D., Nichols, D., Oki, B., Terry, D., "Using collaborative filtering to weave an information

- tapestry". *Communications of the ACM*, 35(12): 61-70, 1992
- [8] Goldberg, K., Roeder, T., Gupta D., Perkins, C., "Eigentaste: a constant time collaborative filtering algorithm". *Information Retrieval*, 4(2): 133-151, 2001
- [9] Good, N., Schafer, J. B., Konstan, J., Borchers, A., Sarwar, B., Herlocker, J., Riedl, J., "Combining collaborative filtering with personal agents for better recommendations". *Proceeding of the 16th National Conference on Artificial Intelligence*, pp. 439-446, 1999
- [10] Han, J., Fu, Y., "Mining multiple-level association rules in large databases". *Knowledge and Data Engineering*, 11(5): 798-804, 1999
- [11] Han, J., Kamber, M., *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2000
- [12] Herlocker, J., Konstan, J., Riedl, J., "An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms". *Information Retrieval*, 5, 287-310, 2002
- [13] Jester Data Download Page: <http://www.ieor.berkeley.edu/~goldberg/jester-data/>
- [14] Kim, B. M., Li, Q., Kim, J. W., Kim, J., "A new collaborative recommender system addressing three problems". *Proceeding of the 8th Pacific Rim Intl. Conference on Artificial Intelligence*, 2004
- [15] Kim, C., Kim, J., "A recommendation algorithm using multi-level association rules". *Proceeding of the IEEE/WIC Intl. Conference on Web Intelligence*, pp. 524-527, 2003
- [16] Konstan J., Miller B., Maltz, D. Herlocker J., Gordon, L., Riedl, J., "Applying collaborative filtering to Usenet news". *Communications of the ACM*. 40(3): 77-87, 1997
- [17] Li, Q., Kim, B. M., "Clustering app. roach for hybrid recommender system". *Proceeding of the IEEE/WIC Intl. Conference on Web Intelligence (WI'03)*, pp. 33-38, 2003
- [18] Lin, W., Alvarez, S. A., Ruiz, C., "Efficient adaptive-supp. ort association rule mining for recommender systems". *Data Mining and Knowledge Discovery*, 6(1): 83-105, 2002
- [19] Linden, G., Smith, B., York, J., "Amazon.com recommendations: item-to-item collaborative filtering". *Industry Report, IEEE*, pp. 76-80, 2003
- [20] Maltz, D., Ehrlich, K., "Pointing the way: active collaborative filtering". *Proceeding of ACM CHI '95 Conference on Human Factors in Computing Systems*, ACM Press, pp. 202-209, 1995
- [21] Miyahara, K., Pazzani, M. J., "Collaborative filtering with the simple Bayesian classifier". *Proceeding of the 6th Pacific Rim Intl. Conference on Artificial Intelligence*, pp. 679-689, 2000
- [22] NutKing: <http://itr.itc.it:8080/dev/jsp/index.jsp>
- [23] O'Connor, M., Herlocker, J., "Clustering items for collaborative filtering". *Recommender Systems Workshop*, 1999
- [24] Open Tourism Consortium Knowledge Repository: <http://www.opentourism.org/wiki/wiki.phtml>
- [25] Pennock, D.M., Horvitz, E., "Collaborative filtering by personality diagnosis: a hybrid memory- and model-based app. roach". *Proceeding of the 16th Conference on Uncertainty in Artificial Intelligence*, pp. 473-480, 2000
- [26] Poon, A., *Tourism, Technology and Competitive Strategies*. CAB Intl., Wallingford, U.K, 1993
- [27] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J., "GroupLens: an open architecture for collaborative filtering of netnews". *Proceeding of CSCW '94*, pp. 175-186., 1994
- [28] Ricci, F., "Travel recommender systems". *IEEE Intelligent Systems*, 17 (6): 55-57, 2002
- [29] Sarwar, B.M., Karypis, G., Konstan, J., Riedl, J., "Item-based collaborative filtering recommendation algorithms". *Proceeding of WWW10*, pp. 285-295, 2001
- [30] Sarwar, B.M., Konstan, J. Borchers, A., Herlocker, J., Miller, B., Riedl, J., "Using filtering agents to improve prediction quality in the GroupLens Research collaborative filtering system". *Proceeding of CSCW '98*, pp. 345-354, 1998
- [31] Shardanand, U., Maes, P., "Social information filtering: algorithms for automating 'word of mouth'". *Proceeding of CHI '95*, pp. 210-217, 1995
- [32] Ski Europe: <http://www.ski-europe.com/>
- [33] Stabb, S., Werther, H., "Intelligent systems for tourism". *IEEE Intelligent Systems*, 17(6): 53-55, 2002
- [34] Watson, I., *App. lying Case-Based Reasoning: Techniques for Enterprise Systems*. Morgan Kaufmann Publishers, Inc., 1997

Web Service Based Universal Management of Workflow Resources

Jinyoung Jang, Yongsun Choi

Dept. of Systems Management & Engineering, Inje University, Kimhae, 621-749, Korea
 pongsor2@hotmail.com, yschoi@inje.ac.kr

ABSTRACT

Implementing business process solutions in the way of Web service is being positioned in the center of workflow management. However, there is no robust standard to expose and access workflow resources by Web service interfaces. In this paper, we propose a web service based workflow resource management framework named Universal Resource Management Framework (URMF) with declarations of web service interfaces and interaction protocols among them. We also introduce a substitutive workflow interface model employing Web services and URMF. Finally, a prototype implementation on model of URMF with J2EE platform is also introduced.

Keywords: Web services, workflow, resource management, organizational resource, resource resolution

1. INTRODUCTION

Workflow management is getting more attention as a way of fostering efficiency potentials by eliminating transport and wait times between activities and providing a detailed level of control over the assignment of work to process participants ([25], [26]). Workflow implementation inside the boundaries of single organization is steadily increasing, however the use of workflow management for the coordination of inter-enterprise processes is still at a very early stage ([5], [31]). Recently, Web services technology, that carries interactions through the exchange of XML messages, provides an alternative approach to implementing interoperable business process solutions in an efficient and standard way ([8], [19]). Application integration through language-neutral and environment-neutral Web services model yields flexible, loosely coupled, and highly extensible business systems.

Current web service composition languages focus mainly on process control perspective but not much on organizational perspective. They lack seriously in managing the organizational elements, such as organization chart, roles, and worklists for human or non-human agents [1], which play important roles in executing business processes. For the successful implementation of workflows in organizations, it is required to enhance web services in supporting organizational resources management.

In this paper, we present a web service based workflow resource management framework named *Universal Resource Management Framework (URMF)* that declares Web service interfaces and the interaction protocols among them. We also introduce a substitutive workflow interface model employing Web services and URMF. Finally, a prototype implementation model of URMF with J2EE platform is also introduced. This paper is structured as follows. Section 2 introduces the issues of organizational resource management in workflow and the related works. Section 3 describes the advantageous features Web services as a resource

management facility. Section 4 describes our Web service based framework for the universal management of workflow resources, especially of organizational resources. An example implementation model of our framework with J2EE platform is described in section 5. Section 6 concludes the paper.

2. ORGANIZATIONAL RESOURCE MANAGEMENT IN WORKFLOW

Workflow management technology supports the execution of business processes through the automated coordination of tasks, data, application logic, and workflow participants (resources). Ideally, the workflow enactment service should be neutral to the organization and resource aspects. Building organization specific services independently from other perspectives, like of process control, can support the needs of wide variety of organizations and resources [33].

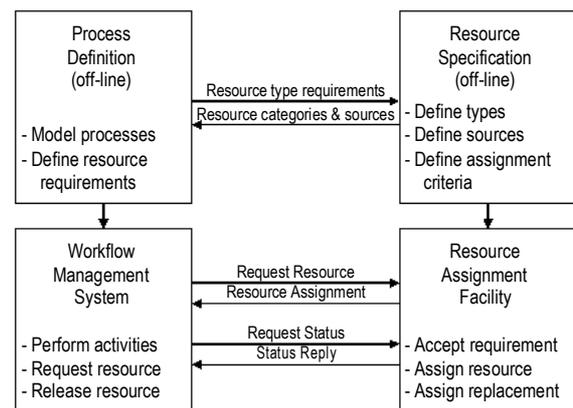


Fig. 1. Separation of resource management perspectives in workflow [21]

Fig. 1 illustrates the separation of resource management perspectives in workflow. Division between a process model on one side and a resource model on the other side fosters the separate evolution of both models. It gives robustness to workflow models independent of changes in the organizational structure. Separate

resource model shared by several workflow engines reduces administrative overhead and prevents possible redundancies [33]. Existing workflow management systems have difficulty in distributed worklist management resulted from the architecture of integrated worklists with the enactment service ([21], [22], [30]).

Resource management framework in workflow can be classified into three categories ([33], [34]). Fig. 2 shows the relationship among these categories and related studies. Main focus of each approach can be summarized as follows:

- *Organizational meta model*. Studies under this category suggest the meta models of organizational structure to be used for the mapping of an existing organization. These meta models is moving to be more generic and comprehensive to cover organizational objects such as organizational positions, position types, tasks, or units. Exposing too rich models tend to cause too tightly coupled and highly dependent architecture of resulting systems ([32], [33]).
- *Resource Resolution*. Upon the instantiation of a workflow activity, the workflow enactment service places work items on the work lists of qualified performers [30]. Resource resolution handles on the strategy and the mechanism for work allocation among competing process participants, utilizing role hierarchy, delegation rules, and various organizational constraints like “binding of duties” and “separation of duties”. Formal resource resolution languages like Resource Query Language (RQL) [16] are employed to provide a ‘view’ to access the inner organization model, in finding out eligible resource, without the direct exposure of the entire organizational model.
- *Resource Assignment / Coordination*. This layer provides a mechanism of how to reserve, cancel, or queue workitems into the worklists of selected workflow resources. Services in this layer may be exposed by APIs such as CORBA IDL.

<p style="text-align: center;">Resource Assignment / Coordination</p> <ul style="list-style-type: none"> • <i>Work Activity Coordination System</i> [14] • <i>Enterprise Workflow Management Framework</i> [12] • <i>Resource Assignment Interface</i> [21] • <i>Workflow Resource Management Facility</i> [11]
<p style="text-align: center;">Resource Resolution</p> <ul style="list-style-type: none"> • <i>Resource Policy Model</i> ([6], [7]) • <i>Resource Query Language</i> [16]
<p style="text-align: center;">Organizational Meta Model</p> <ul style="list-style-type: none"> • <i>OMM Organization and Role Model</i> [9] • <i>Organization and Resource Model (ORM)</i> ([23], [24]) • <i>Generic Organization Model</i> ([6], [7]) • <i>A Generic Resource Meta Model</i> [33]

Fig. 2. Issues and related studies of organizational resource management in workflow

However, these works on organizational resource management in workflow are limited in handling the issues of interoperability and heterogeneity, which can

be overcome by the web service technologies. In next section, we will discuss how resource management framework can be modeled and implemented with the web services.

3. WEB SERVICES AND RESOURCE MANAGEMENT

Existing workflow resource management approaches are much dependent on the underlying models or technologies, resulting tightly coupled architectures. They have difficulty in allowing the flexible and coordinated interaction patterns, which can be provided by the web service architecture. Although there exist many distributed software platforms that could be useful to implement workflow resource facility, Web services technologies are getting more industry momentum with many advanced features, like language and platform independence, interoperability through wire level standards, and the ability to expose interfaces but hiding implementation [15].

Today, the most prevalent use of Web services is to carry out business processes and transactions, inside or across enterprises, by applying web service composition languages like BPEL4WS. However, there exist few studies of employing web services for the management of organizational resources in workflow. Recently, Aalst et al. [1] introduced their work on how to expose an organizational model in an XML DTD for the exchange of information in cross-organizational processes. But, their approach still bears tightly coupled architecture.

Managing organizational resources by Web services can be a sound alternative by utilizing tools and standards of Web services. Two of the existing works that can be employed to adapt workflow resources to Web services are introduced.

Intermediary server

An intermediary server is a component that lies between the service client (subscriber) and the service provider (publisher). It basically intercepts the request from the service client, provides the service of designated function, and forwards the request to the service provider. Similarly, it intercepts the response from the service provider and forwards it to the service client. Intermediary servers can provide value-added services, like authentication, quality of service, management service, or aggregation services. In our approach, we employ intermediary servers as to provide a uniform service of combining different types of organizational resources into a single unified interface.

Web Service Management Framework

Web Services Management Framework (WSMF) is proposed for the management of various IT resources, including Web services themselves, through Web services [15]. To encounter various management issues of more complex IT resources, an extensible interface

has been defined utilizing Web services standards and description language. However, WSMF focuses on the managerial aspects of IT resources, such as network management, system management, or database management, which are far from the management issues of organization resources, such as work reservation, work coordination, or organization charts.

4. UNIVERSAL RESOURCE MANAGEMENT FRAMEWORK

Now, we introduce a web service based workflow resource management framework named *Universal Resource Management Framework (URMF)*. URMF provides several web service interfaces that support in browsing organization chart and belonging resources, interacting with human and non-human resources, and the dynamic selection of proper resources based on constraints.

4.1. Design principles of URMF

The design principles of URMF, differentiated from the traditional distributed computing, are based on the concept of the service-oriented architecture (SOA) [20] to make the workflow resource management facility conform to the Web service environment:

- **Standardized.** The URMF architecture conforms to the web service standards (SOAP, WSDL, and UDDI) including the web service based process definition languages such as BPEL4WS. Resources declared and managed in URMF are universally discovered, bound and executed in web service environment.
- **Loosely coupled by coarse-grained interfaces.** Distributed object based systems are moving toward coarser-grained interfaces, which act as distributed façades, for less communication with each other across the network and loosely coupled implementation [20]. URMF provides coarse-grained web service interfaces, by employing generic organization model and resolving resources without the direct exposure of the entire organizational model. These features allow the flexible enrichment of organizational meta models, the employment of diverse underlying technologies, e.g. LDAP (Light-weight Directory Access Protocol) or SQL, and the flexible coordination of the services.
- **Asynchronous work coordination.** Tasks in business processes, especially those performed by human resources, need to be queued waiting to be processed. URMF supports the message calls for work order asynchronously as well as synchronously by the event-driven mechanism, to be explained more detail in sections 4.3, 5.2 and 5.3.
- **Dynamic selection of resources.** URMF provides web service interfaces, for the dynamic selection of resources, which supports the various organizational constraints represented in Resource Query Language (RQL) as well as browsing organizational structure, to be explained more detail in Sections 4.3 and 5.5.
- **Discoverable.** All Web services, inside or outside of

an organization, are discoverable as a benefit of SOA, either at design time or at run time. Organizational resources conforming to URMF can be discovered, at any time required regardless of the location, to be explained more detail in Sections 4.3 and 5.6.

4.2. Entities of URMF and their interactions

Fig. 3 shows the four main entities of URMF, i.e., *WSWE (Web Services-based Workflow Engine)*, *URB (Universal Resource Binder)*, *OWQS (Organization Work Queue Service)*, and *AWQS (Agent Work Queue Service)*.

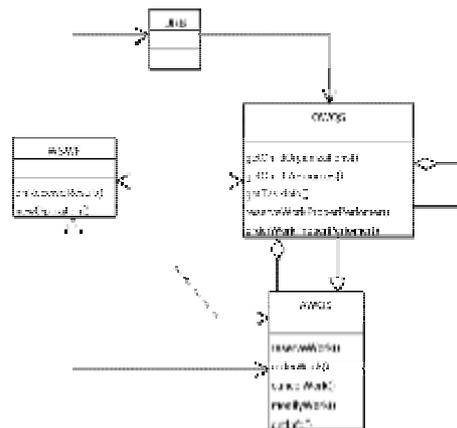


Fig. 3. Entities of URMF and their interactions

In URMF, every organizational workflow resource is represented as a web service, called *Work Queue Service (WQS)*, and acts as a service provider. Each *AWQS* is for a human or non-human resource [30] and takes the work requests of service requestors. Each *OWQS* is for an organization unit and acts as a hub or router for *AWQS*. On the other hand, *WSWE* or *URB* request the services provided by *WQS*. *WSWE* refers to the workflow engine capable of interacting with web services. *URB* refers to a GUI tool, intermediating *WSWE* and *WQS*, and supports for browsing organizational resources during process design or instantiation.

AWQS represents each workflow resource, such as a human agent, software, a machine, or a place. *OWQS* inherits the interfaces of *AWQS*, capable of acting as *AWQS*, and references one or more child *AWQS* or *OWQS*. This self-contained structure allows *OWQS* to cover any set of resources like companies, branches, departments, or teams. This composite pattern [13] allows each service unit to be declared for any number of *WQS*'s and lets URMF be applied to the organizations of any scale either in size or in complexity of policies. This advanced feature additionally allows *OWQS* to provide the functionality as an intermediary server. For example, *OWQS* can delegate or distribute work items to the appropriate child resources under its policy of delegation rules, access control, or security, etc. Furthermore, *OWQS* provides the RQL interfaces so that the service requestors can get the endpoint of the desired resource without browsing the whole organization charts inside. In this way, should-be secure

objects, like level or delegation, are encapsulated. These structural features of URMF entities support the design criteria of loosely coupled and coarse-grained interface.

4.3. Services of URMF

With the design principles and the entity model, we have declared five categories of services for URMF. Specific Web service descriptions of URMF entities, classified into these categories, are provided in appendix A. In addition, an example process definition using these services in BPEL4WS is provided in appendix B.

- **Organization Chart Service** provides several subordinate services to access organizational structure and resources. This service is mostly invoked at the design time or when a process is instantiated, by a process designer or an administration tool with the functionality of URB, for the actual binding of endpoints of target resources.

- **Resolution Service** enables service requestors to select resources dynamically with the Resource Query Language (RQL). This service returns the endpoint of proper AWQS according to the input query, which can be dynamically built according to the flow of process instance. XML-based string manipulation expressions, e.g. 'XPath' [10], can be used to build changeable RQL statements automatically.

- **Work Coordination Service** enables service requestors to reserve, order, cancel, and modify work items to the target resources. It also provides the runtime-state information of started tasks and reservations. The service requestor may call the AWQS direct through the endpoints of each workflow resource. But this direct access to resources may cause security problems and is restrictively to be used only inside an enterprise. On the other hand, the service requestor may call the OWQS and the corresponding OWQS delegates the work order to the final AWQS. This delegated work coordination provides better security management in the form of an intermediary server.

- **Event Service** enables service requestors to subscribe and listen to the events of interested resources, tasks, or changes in the organization. This service allows the service requestors to browse the list of events of the objects and subscribe to the event with the event key. Task related events, like a work is reserved or queued, are notified to the corresponding service requestor without subscription.

- **Universal Discovery Support** URMF requires each WQS instance to have its own address of web service endpoint URI. With this universal addressing, each resource in URMF can be registered to and discovered by the UDDI directory. By utilizing some taxonomy of role keys at WQS registration, resource searching can be further enhanced.

4.4. Web service based workflow interface model with URMF

Fig. 4 illustrates the substitutive workflow interface

model, modified from [30], employing URMF and underlying web services architecture. Each interface 1 through 5 is mapped with corresponding Web services standards. Each component in the interface model takes advantage of the Web services infrastructure such as universal discovery, transaction management, or security management. In this way, the URMF architecture offers workflow resource management capabilities in a standard way according to the evolving Web services specifications.

Web service composition languages (WSCL), like BPEL4WS [3], conform to the web service architecture and can be extended to describe resources by URMF. Interface 1 in our model (*IF1* in Fig. 4) focuses on the import and export of process definitions in WSCLs. In our interface model, web service interface AWQS substitute worklists of each end-user (*IF2*) or the non-human resources like software agents (*IF3*). Sharing URMF-conformed resources, with other workflow engines, provides a way of enhancing interoperability (*IF4*).

Finally, the Universal Role Binder (URB), a GUI tool of URMF, interacts with process designer, administration/monitoring tool, and work queue services. URB supports end-users in searching and binding the resources in the way of web service as shown in Fig. 4.

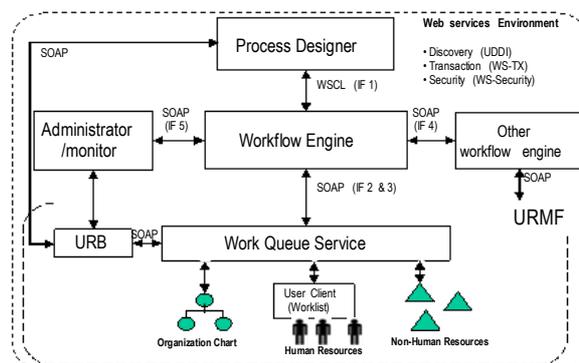


Fig. 4. Web Service based workflow interface model with URMF

As illustrated in Fig. 4, URMF complements the resource management facility of Web services environment. Since the interfaces and the tool declared in URMF extract and provide the resource management functionalities, the efforts of design and implementation on this part are relieved when to expose workflow resources as web services. In next section, a prototype implementation model of exposing workflow resources as web services with URMF is introduced.

5. PROTOTYPE IMPLEMENTATION MODEL

In this section, we show an example implementation model of URMF and how the web service requests are processed by the implemented system. Two suggestions for the realization of universal discovery support in

URMF are also presented.

Fig. 5 illustrates the system architecture of the prototype implementation model. For better management of reliability, security, and scalability, we adopted one of the most concurrent technologies, i.e. J2EE platform [28]. The interfaces and components to support URMF are managed by the application server, and intermediate the service requestors and the workflow resources of service providers. Service requests in the form of web service invocations are received at the two interfaces (AWQS and OWQS), processed at the corresponding components, and provided with low-level access to the adequate resources. More detailed descriptions of the major components in the implemented model are provided.

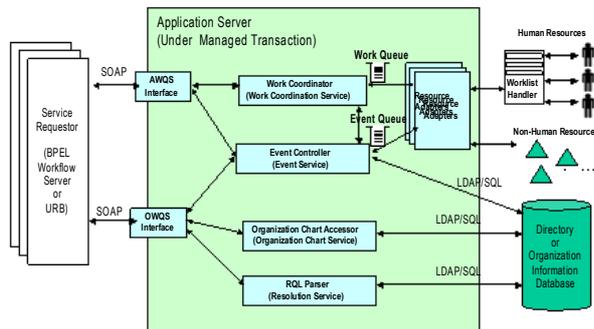


Fig. 5. System architecture of the prototype URMF Implementation

• Work Coordinator

Work Coordinator processes the requests of reserving, queuing, canceling and modifying work items from AWQS interfaces. For the reliable management of work requests, MQM (message queuing middleware) and managed-transaction of J2EE platform are adopted ([27], [28]). Each work order or work reservation is queued into the message queue named 'WorkQueue' and is ensured to be properly processed. The *resource adapters* let the work requests be uniformly processed at the heterogeneous system resources. *Worklist handler* provides a user interface for end users to handle workitems with adequate tools.

• Event Controller

Event Controller manages the subscription and notification of events. Event Controller listens to the events of interested resources, tasks, or organization changes. Each event fired from other components is queued into the message queue named 'EventQueue'. Event Controller distributes them to the subscribed service requestors (URB or WSWE). Task related events, e.g., 'onResult' or 'onReservationResult' of appendix A, are notified to the corresponding service requestor even without subscription.

• Organization Chart Accessor

Organization Chart Accessor processes the requests for accessing organizational structure or belonging resources on behalf of the *organization chart services* of

WQS interfaces. Each request is processed by accessing the directory services through LDAP [18] or the organization information database through SQL.

• RQL Parser

RQL Parser executes the requests for the endpoint of eligible resource (WQS) on behalf of *resource resolution services* of OWQS. The inquiry RQL statements are parsed and dispatched through LDAP or SQL.

• Universal discovery support

Web services are discoverable by their own nature. To be discovered by UDDI, WQS of each resource needs unique address and categories. Two suggestions to handle these issues are described.

WQS endpoint (URI) addressing: WQS of each resource in an organization needs a unique address, i.e. endpoint URI. Fig. 6 shows the suggested endpoint addressing policy for WQS. It is similar with e-mail addressing but contains more organizational information, i.e. names of department, branch or company as well as and user id.

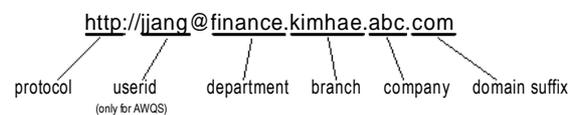


Fig. 6. Suggested WQS endpoint (URI) addressing

Role-based Taxonomy: Mutually aware taxonomies are needed to search for adequate web services through UDDI. We suggest employing role-based taxonomy to be utilized in searching for adequate WQS of a resource. Utilizing role-based taxonomy, WQS of adequate human resources with required role or skill could be searched from UDDI servers.

Fig. 7 illustrates the suggested mechanism of universal discovery support, which allows each organization or person to advertise itself, as a resource, over the Web service world.

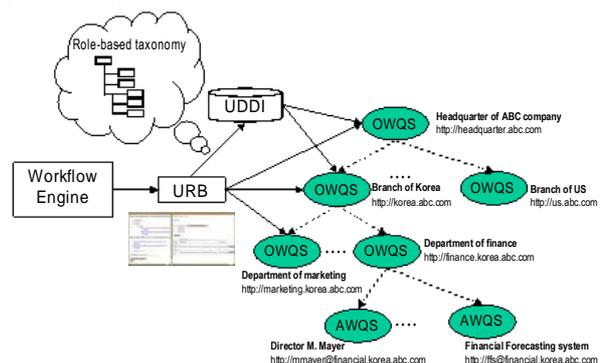


Fig. 7. Suggested mechanism of universal discovery support

6. CONCLUDING REMARKS

Organizational elements, such as organization chart, roles, and worklists for human or non-human agents, play important roles in executing business processes [1]. For the successful realization of web service based workflow management, a sound management of these elements should be provided. However, there is no

robust standard to expose and access workflow resources by Web service interfaces. In this paper, we proposed a web service based workflow resource management framework named *Universal Resource Management Framework* (URMF) with declarations of web service interfaces and interaction protocols among them. We also introduced a substitutive workflow interface model employing Web services and URMF, where each component of the model can take advantage of Web services infrastructure. To validate the viability of the suggested framework, a prototype implementation model of URMF with J2EE platform is also introduced. In this way, management of business processes composed of Web services is empowered great with the support of URMF in managing organizational resource.

Our framework is fully based on the service-oriented architecture (SOA) [20]. The resources, especially human resources, conforming to URMF are universally interoperable and discoverable as web services, and can be flexibly coordinated in the way of Web services standards. Loosely coupled and coarse-grained URMF interfaces, by the generic organization model and resolving resources with the encapsulation of the entire organizational model, features minimal technical requirements for the implementation.

Some of important issues related with the Web services environment are not dealt in this paper and left as further studies. First, Non-authorized work request or access to the organization should be prohibited. Since URMF is based on Web services, Web services security technologies may leverage the security of our framework. The SSL certificates with HTTPS or WS-Security [17] are examples of those technologies. Some access control mechanism may be applied to the URMF at the interface or operation level. For instance, OWQS can act as an intermediary server for security and provide an access control mechanism based on the service requestor or information requested. Second, an adequate business transaction mechanism to handle long-lived activities should be supported with URMF. To minimize latency of access, for other potential users, to the resources used by such activities, the results of interim operations need to be released before they are completed. URMF needs to be enhanced by applying such frameworks for transaction management, like WS-Transaction [4] or WS-CAF [2], which coordinates the execution of distributed operations in a Web services environment.

ACKNOWLEDGEMENT

This work was supported by the 2003 Inje University research grant.

REFERENCES

[1] Aalst, W.M.P. van der, A. Kumar, and H.M.W. Verbeek, "Organizational Modeling in UML and XML in the context of

Workflow Systems", *18th Annual ACM Symposium on Applied Computing*, ACM Press, 2003, pp. 603-608.

[2] Arjuna Technologies Ltd., Fujitsu Limited, IONA Technologies Ltd., Oracle Corporation, and Sun Microsystems, Inc., "Web Services Composite Application Framework (WS-CAF). Version 1.0", July 2003.

[3] BEA Systems, IBM Corporation, and Microsoft Corporation, Inc., "Business Process Execution Language for Web Services, Version 1.1", <http://www.ibm.com/developerworks/library/ws-bpel/>, May 2003.

[4] BEA Systems, IBM Corporation, and Microsoft Corporation, Inc., "Web Services Transaction (WS-Transaction)", <http://www.ibm.com/developerworks/library/ws-transpec/>, August 2002.

[5] Bussler, C., "Enterprise-wide workflow management", *IEEE Concurrency*, Vol. 7, Issue 3, 1999, pp. 32-43.

[6] Bussler, C. and S. Jablonski, "Policy resolution for workflow management systems", *28th Hawaii Int. Conf. on System Sciences*, Vol. 4, 1995, pp. 831-840.

[7] Bussler, C., "Analysis of the Organization Modeling Capability of Workflow-Management-Systems", *the PRIISM '96 Conf.*, January 1996.

[8] Chappell, D. A. and T. Jewell, *Java Web Services*, O'Reilly & Associates, Inc., March 2002.

[9] Cheng, E. C., "An Object-Oriented Organizational Model to Support Dynamic Role based Access Control in Electronic Commerce Applications", *the 32nd Annual Hawaii Int. Conf. on System Sciences*, 1999.

[10] Clark, J. and S. DeRose (ed.) "XML Path Language (XPath) Version 1.0", W3C, November 1999, available at <http://www.w3.org/TR/xpath>

[11] Cummins, F. A., "Workflow Resource Management Facility Request for Proposal (Draft)", *OMG Document bom/99-04-01*, Framingham 1999.

[12] Du, Weimin, J. Davis, Y.-N. Huang, M.-C. Shan, "Enterprise Workflow Resource Management", *HP Tech. Report*, HPL-1999-8, 1999.

[13] Gamma, E., R. Helm, R. Johnson, and J. Vlissides, *Design Patterns—Elements of Reusable Object-Oriented Software*, Addison-Wesley, 1995.

[14] Halliday, J.J., S.K. Shrivastava, and S.M. Wheeler, "Implementing support for work activity coordination within a distributed workflow system", *3rd Int. Conf. on Enterprise Distributed Object Computing Conference*, 1999, pp. 116-123.

[15] Hewlett-Packard Development Company, "Web Services Management Framework – Overview Version 2.0", July 2003.

[16] Huang, Y.-N. and M.-C. Shan, "Policies in a Resource Manager of Workflow Systems: Modeling, Enforcement and Management", *HP Tech. Report*, HPL-98-156, 1998.

[17] IBM Corporation, Microsoft Corporation, and VeriSign, Inc., "Web Services Security (WS-Security) Version 1.0 05", <http://www.ibm.com/developerworks/library/ws-secure/>, 2002.

[18] LDAP, *Understanding LDAP*, 1998, <http://www.redbooks.ibm.com/redbooks/pdfs/sg244986.pdf>

[19] Leymann, F., D. Roller, and M.-T. Schmidt, "Web services and business process management", *IBM Systems Journal*, Vol. 41, No. 2, 2002, pp. 198-211.

[20] McGovern, J., S. Tyagi, M. Stevens, and S. Mathew, *Java Web Services Architecture*, Morgan Kaufmann, 2003.

[21] Object Management Group (OMG), "Workflow Resource Assignment Interfaces (RAI) Request For Proposal", *OMG Document bom/2000-01-03*

[22] Paul, S., E. Park and J. Chaar, "RainMan: A Workflow System for the Internet", *Proc. of USENIX Symp. on Internet Technologies and Systems*, November 1997.

[23] Rupiotta, W., "Organization and Role Models for Workflow Processes", In Lawrence, P.(Ed.) *The Workflow Management Coalition Handbook*, Winchester 1997, pp. 165-172.
 [24] Rupiotta, W., "Organizational Models for Cooperative Office Applications", In D. Karagiannis (Ed.) *Database and Expert Systems Applications. the 5th Int. Conf. DEXA '94*, Springer Publishers, September 1994.
 [25] Sheth, A. P., W. van der Aalst, and I.B. Arpinar, "Processes driving the networked economy", *IEEE Concurrency*, Vol. 7, Issue 3, 1999, pp. 18-31.
 [26] Stohr, E. A. and J. L. Zhao, "Workflow automation: Overview and research issues", *Information Systems Frontiers*, 3(3):281-296 (2001).
 [27] Sun Microsystems, Inc., "Java Message Service, v1.0.2b", August 2001, <http://java.sun.com>
 [28] Sun Microsystems, Inc., "JavaTM 2 Platform Enterprise Edition Specification, v1.4", April 2003, <http://java.sun.com>

[29] Swenson, K. D., "Trouble Ticket Workflow Scenario (Version 2)", *OMG Document bom/98-02-09*.
 [30] Workflow Management Coalition, *The Workflow Reference Model: Document Number TC00-1003*, 1995, available at <http://www.wfmc.org>.
 [31] zur Muehlen, M., "A Framework for XML-based Workflow Interoperability - The AFRICA Project", *Americas Conference on Information Systems (AMCIS 2000)*, pp. 13-18.
 [32] zur Muehlen, M., "Evaluation of Workflow Management Systems Using Meta Models", In the 32nd Annual Hawaii Int. Conf. on System Sciences, 1999.
 [33] zur Muehlen, M., "Resource modeling in workflow applications", *Proceedings of the 1999 Workflow Management Conference*, 1999, pp. 137-153.
 [34] zur Muehlen, M., "Organizational Management in Workflow Applications", *Information Technology and Management Journal*, Kluwer Academic Publishers, Vol. 5, No. 3, 2004.

APPENDIX A. SERVICE DESCRIPTIONS OF URMF ENTITIES

Table A1. Service description of AWQS

Service Type	Service Name	Input & Output
Work Coordination Service	reserveWork	I: Role Key, Reply Endpoint, Expected Order Date, Due Date; O: Task (Reservation) Key
	orderReservedWork	I: Task (Reservation) Key, Parameter (any type); O: Task Key
	orderWork	I: Role Key, Reply Endpoint, Due Date, Parameter (any type); O: Task Key
	cancelWork	I: Task Key, Reason; O: None
	modifyWork	I: Task Key, Due Date, Parameter (any type); O: None
	modifyReservedWork	I: Task Key, Expected Order Date, Due Date; O: None
	getTaskInfo	I: Task Key; O: Task Information
Event Service	subscribeEvent	I: Event Type, Reply Endpoint, Duration; O: Event Key
Organization Chart Service	getInfo	I: None; O: Resource information

Table A2. Service description of OWQS*

Service Type	Service Name	Input & Output
Organization Chart Service	getChildOrganizations	I: None; O: Endpoints of child OWQS's
	getParentOrganization	I: None; O: Endpoint of the parent OWQS
	getAgents	I: None; O: Endpoints of child AWQS's
	getRoleKeys	I: None; O: List of all the role keys the organization resources can play
Resolution Service	getProperPerformer	I: RQL Statement; O: Endpoint of the resolved WQS
Resolution Service + Work Coordination Service	reserveWorkProperPerformer	I: RQL Statement, Reply Endpoint, Expected Order Date, Due Date; O: Task (Reservation) Key
	orderWorkProperPerformer	I: RQL Statement, Allocation Policy, Reply Endpoint, Due Date, Parameter (any type); O: Task Key

* Basically, OWQS inherits all the operations of AWQS as described in 4.2.

Table A3. Service description of WSWE

Service Type	Service Name	Input & Output
Event Service	onResult	I: Task Key, Result of work (any type); O: None
	onReservationResult	I: Task (Reservation) Key, Result of reservation; O: None
	onEvent	I: Event Type, Event Context; O: None

APPENDIX B. AN EXAMPLE BPEL4WS CODE EMPLOYING URMF

The following BPEL4WS code illustrates how the human resources can be dynamically selected and bound with URMF. The code describes parts of the trouble ticket process¹, where each problem instance is recorded with human

¹ The trouble ticket scenario by OMG [29] is provided for the purpose of evaluating workflow facilities in the perspective of human intervention.

intervention. The process is initiated by the former “receive” activity, which receives the endpoint of the OWQS of the organization in charge of the given problem instance. Based on the request, the “assign” activity builds a RQL statement of inquiring eligible resource and set parameters for the invocation of corresponding WQS. The “invoke activity” calls the interface (“**orderWorkProperPerformer**” in this example) of the WQS for the resolution and assignment of this workitem. Actual selection and binding of the proper human resource (i.e. queuing the workitem in the corresponding worklist) is executed by the *resource resolution service* and the *work coordination service* of WQS. The latter “receive” activity waits for the result performed by the AWQS of actually assigned human resource.

```

<!-- declaration of message types, variables, and correlation sets will be used -->
<message name="TroubleTicketRequestType">
  <part name="originatorOWQSServiceReference" type="wsa:EndpointReference"/>
</message>
<containers>
  <container name="troubleTicketRequest" messageType="tns:troubleTicketRequest"/>
  <container name="OWQSRequest" messageType="owqs:WorkRequest"/>
  <container name="taskKey" messageType="wqs:TaskKey"/>
  <container name="problem" messageType="wswe:WorkRequest"/>
</containers>
<correlationSets>
  <correlationSet name="TaskIdentification" properties="tns:TaskKey"/>
</correlationSets>
<!-- the process will be initiated with the originating organization's service reference (OWQS endpoint) -->
<receive partner="originator" portType="tns:TroubleTicketPT"
  operation="initiate" container="troubleTicketRequest"
  createInstance="yes" name="TroubleTicketReceive"/>
<!-- build parameters for invoking AWQS -->
<assign>
  <!-- provide a RQL for resolving the proper performer for this activity -->
  <copy>
    <from expression="require humanworker for TroubleTicket:RecordingProblem where branch.location=KR"/>
    <to container="OWQSRequest" part="rqlStatement"/>
  </copy>
  <!-- provide the reply endpoint for the result of the activity -->
  <copy>
    <from expression="http://korea.abc.com/bp/troubleticket"/>
    <to container="OWQSRequest" part="replyEndpoint"/>
  </copy>
  <!-- copy the service(endpoint) reference of originator's AWQS --> <copy>
    <from container="originatorOWQSServiceReference"/>
    <to partner="recorder"/>
  </copy>
</assign>
<!-- queuing a workitem into the performer's worklist -->
<invoke name="RecordingProblem"
  partner="recorder"
  portType="owqs:OrganizationWQS"
  operation="orderWorkProperPerformer"
  inputContainer="OWQSRequest"
  outputContainer="taskKey">
  <!-- correlation with the Task key that was returned as result of invocation -->
  <correlations>
    <correlation set="TaskIdentification" initiation="yes"/>
  </correlations>
</invoke>
<!-- receive the result of the requested workitem -->
<receive name="ReceiveResultFromRecordingProblem"
  partner="recorder" portType="wswe:WSWE" operation="onResult" container="problem">
  <correlations>
    <correlation set="TaskIdentification" initiation="yes"/>
  </correlations>
</receive>

```

A Comparative Analysis of International and Chinese Electronic Commerce Research

Junyong Xiang, Jaekyeong Kim, Hyouk O

School of Business Administration, Kyung Hee University, Seoul, 130-701, Korea
gysang@yahoo.com, jaek@khu.ac.kr, cuteing@gmail.com

ABSTRACT

Due to the growth of the Internet and e-commerce, both practitioners and researchers are in the midst of a social, business and culture revolution. Internet and e-commerce related research has been developed and grown up by United States, but China has become one of the most exciting research areas. This literature review consists of 1044 journal articles published between 1993 and 2003 in fourteen International and Chinese journals. The articles are classified by a scheme that consists of four main categories: application areas, technological issues, support and implementation and others. Based on the classification and analysis of e-commerce related researches, we present the current state of International and Chinese research and discuss the differences between them.

Keywords: *e-commerce, classification of e-commerce, literature review, comparative research*

1. INTRODUCTION

With the development of Globalization and IT, business and organization are radically changing. In contrast with the weak performance of several key developed and developing economies in recent years, e-commerce have continued to grow at a fast pace. The United States and European countries leading in IT and e-commerce have experienced impressive GDP growth since 1995. This output expansion has been characterized by acceleration in productivity growth, very low unemployment rates, low inflation rates and a reduction of fiscal deficits. With the booming Chinese economy, WTO accession and 2008 Beijing Olympics, China's e-commerce is also experiencing a fast development period and is becoming a new driving force for the economic development. Therefore, it is very meaningful and interesting to compare the state of International and Chinese research on e-commerce.

Ngai and Wat [3] introduced a classification framework of topics in e-commerce that consists of four main categories: application areas, technological issues, support and implementation and others. They classified the e-commerce research published in international journals between 1993 and 1999. Their classification framework was eventually adopted in this research.

This paper presents a comprehensive review, classifies the literature of International and Chinese e-commerce research published between 1993 and 2003. The scheme we used represents our views of the e-commerce literature that point out the distinguishing features of this area. A classification of e-commerce research reveals that there has been an approximately exponential growth in the number and area of published articles. The current state and direction of research topics should be of interest to researchers and practitioners.

2. CLASSIFICATION SCHEME OF E-COMMERCE RESEARCH

2.1 Research Paper Collection

According to Zwass [7], the first popular Web browser, NCSA Mosaic was introduced in 1993 and it brought people and businesses to the Web. Therefore, we believe that this year may be considered to be the starting point of e-commerce.

We exclude conference proceeding papers, master's theses, doctoral dissertations, textbooks, and unpublished working papers. We believe that academics and practitioners alike use journals most often for acquiring information and disseminating new findings and represent the highest level of research. Then we analyze articles published in International and Chinese journals between 1993 and 2003.

All journals lie in areas of Management Science and IS. Based on existing researches [4,6], the list of Chinese Social Sciences Citation Index (CSSCI) and Chinese Science Citation Database (CSCD) and a review of most of the journals, fourteen journals are selected.

Seven International journals were selected in this study, namely: Communication of the ACM, Decision Support Systems, Electronic Commerce Research and Applications, Information & Management, International Journal of Electronic Commerce, Journal of Management Information Systems, and Journal of Organizational Computing and Electronic Commerce. Seven Chinese journals were selected in this study, namely: Journal of Management Science, China Soft Science Magazine, Science of Science and Management of S. & T., East China Economic Management, Modernization of Management, Journal of Industrial Engineering and Engineering Management, and Computer Engineering and Applications.

Online databases search were conducted. For International journals' databases, we used ABI/INFORM Global (<http://www.il.proquest.com/proquest>), Science Direct (<http://www.sciencedirect.com>) and the ACM Digital Library (<http://portal.acm.org>). And we used China National Knowledge Infrastructure (CNKI) Digital Library (<http://www.cnki.net>) as Chinese journals databases.

The literature search was based on four descriptors: "EDI", "Electronic Commerce (E-Commerce)", "Electronic Business (E-Business)" and "Internet Commerce"[3] related to e-commerce, we reviewed the full text of each article. As a result of our search, we identified 1044 e-commerce articles. Although this

search was not exhaustive because of some limitations, such as online Database update time, it served as a comprehensive base for an understanding of e-commerce research.

2.2 Classification of Topics in Electronic Commerce

The classification framework was based on the literature review, the nature of e-commerce research, the works of [2,5] and existing classification schemes of IS [1] (Figure 1). The articles were classified into four broad categories: (i) applications, (ii) technological issues, (iii) support and implementation and (iv) others; and each were divided into subcategories.

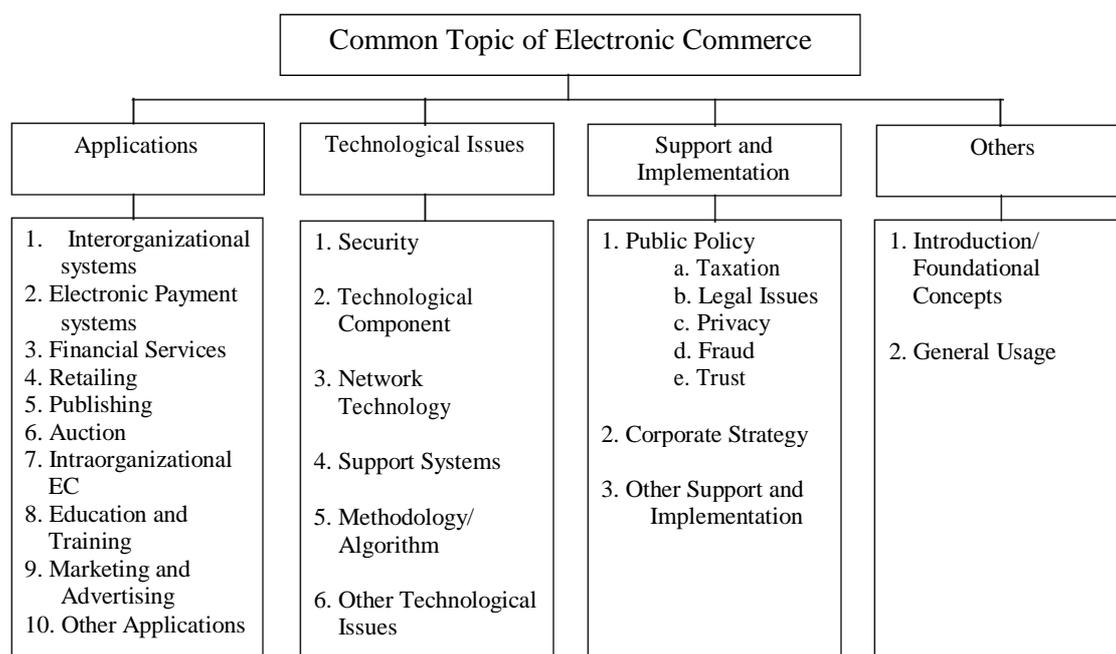


Figure 1. Classification of topics in electronic commerce.

Applications: This category covers literature on e-commerce research related to interorganizational systems (IOS), electronic payment systems (EPS), financial services, retailing, online publishing, auctions, intraorganizational EC, education and training, marketing and advertising, and other applications. IOS are automated information systems shared by more than one organization and allowing information flow across organizational boundaries. They include electronic data interchange (EDI), supply chain management (SCM), electronic funds transfer (EFT), electronic forms, electronic messaging and shared databases. EPS cover systems for online financial exchange between buyers and sellers. For examples, electronic or digital cash, electronic checks, online credit card-based systems, and other emerging financial instruments, such as debit cards, electronic benefit transfer cards, and smart cards. Financial services cover the common financial services provided in e-commerce. Retailing contains various publications associated with topics of retailing in

e-commerce, including online and electronic shopping malls. Online publishing contains publications that relate to electronic delivery of newspapers, magazines, books, news, music, videos, and other digitizable information through the Internet. Auctions contain published literature related to electronic auctions. Intraorganizational EC deals with helping a company to maintain and support internal business processes between individuals, departments and collaborating organizations. Education and training cover published literature on education or training pertaining to on-line education, virtual classrooms and online universities. Marketing and advertising consist of the application of e-commerce to marketing and advertising, including brand-name management, disseminating product catalogs, and sales information and product announcements. Other applications contain publications that discuss applications that are not among those in the above categories, such as Internet gambling, travel and tourism

service, employment placement and the job market, health care, on-line games, matchmaking, etc.

Technological issues: There are six categories pertaining to technological issues: security, technological components, network technology/infrastructure, support systems, algorithm/methodology and other technical issues. Security consists of publications that discuss issues related to data and system security, and individual security. Technological components consist of various Internet technologies, e.g. the common object request broker architecture (COBRA). Furthermore, markup languages, such as hypertext markup language (HTML), and programming languages like JAVA are the common Web software developmental tools. Network technology/infrastructure covers network technology that is at a lower-level, such as early technologies, the local area network (LAN) technology and the wide area network (WAN) technology; network protocols, hypertext transport protocol (HTTP) and transmission control protocol/Internet protocol (TCP/IP); plus network issues, like the shortage of bandwidth, quality of service (QOS). Support systems cover literature on decision support systems, negotiation support system or distributed applications. Algorithm/methodology covers technical algorithms or methodologies that assist, enhance, or improve e-commerce applications. Other technical issues contain publications that discuss other issues such as decision technologies, defense system, and virtual enterprise (organization, marketplace) technology.

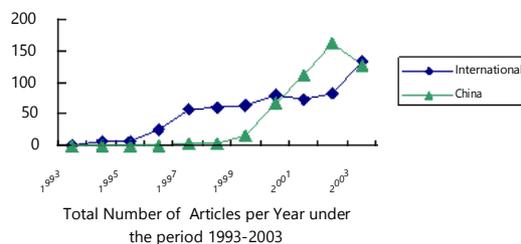
Support and implementation: The literature in this category can mainly be divided into two broad categories: public policy and corporate strategy. And public policy covers taxation, legal issues, privacy, fraud and trust five subcategories. Taxation issues contain the characterization of the transaction, the source of the transaction, the creation of a permanent establishment from the transaction, transfer pricing implications, and tax administration and compliance. Legal issues include privacy, protecting intellectual property, controlling Internet indecency, preventing fraud, establishing a tax framework, controlling gambling, determining jurisdiction, and protecting both sellers and buyers. The major privacy issues are how to protect privacy and whose responsibility it is to do so. Some literature is related to fraudulent practices in e-commerce, such as online auction fraud, Internet stocks fraud and other fraud. The published literature refers to positive trust that is shown in the consistency and assurance between what a trading partner says and actually does. Corporate strategy covers publications dealing with EC strategies/methodology used in building successful implementations. Other support and implementation contain various publications that are not mentioned in the support and implementation categories, such as modeling the structure of brokered electronic markets.

Others: This category contains publications that discuss other aspects of e-commerce, such as a general

introduction to e-commerce, foundational concept of e-commerce, etc.

3. RESULTS

3.1 Distribution of the Year of Publication



Notes: Forecast the Figure from November to December of 2003 of China.

Figure 2. Distribution of articles by year.

The distribution of articles published by year is shown in Figure 2 from 1993 to 2003. Although research outputs appear limited, yet the number of journal articles has increased substantially. International e-commerce research generally activated from 1996, and Chinese research generally activated from 1999. When it comes to 2000s, it does not show apparent difference in amount, yet there are still many gaps in width and depth of the research. It will take some time for China to bridge the gaps.

In 1995 e-commerce was just beginning and its potential was not widely recognized. Since then, B2B and B2C e-commerce has skyrocketed. During 1995 and 1999, American economy kept the 4% increase and about 2% inflation rate per year, and the unemployment rate decreased to 4%. The success of the United States made all countries and international organizations take a serious view of e-commerce. Consequently, the international researches on e-commerce were conducted robustly and kept substantial increase until 2000. The bursting of the new economy bubble late in 2000 made the research suffer a setback temporarily. Yet the increase trend is continuing, and the research will enter a new phase and become more active in recent years.

For Chinese research, the most distinguishing feature is the rapid increase speed. China's first move was the initiation of the Golden Projects. The Government initiated the "Government Online" and "Enterprises Go Online" projects in 1999, and it declared the year of 2000 to be the "Year for Chinese Businesses to Go Online". The national informatization program is given high priority in the 10th Five Year Plan (2001 to 2005). During this period, the information industry is expected to become the pillar industry to spur the national economic growth. Furthermore, the China e-commerce Association was established in June 2000. Therefore, China's e-commerce and research develop at a rapid pace while other countries temporarily suffered

stagnation of development. The research reached the peak in 2002 because of the special issue (2002, Vol.4) on e-commerce of China Soft Science Magazine.

3.2 Distribution of Articles by Topic

The distribution of articles by topic is shown in Figure 3. It appears that the most heavily published research area is in the “Technological Issues” of EC, all have more than 30% in this area (International—196 articles, 33%; China—166 articles 36%). It also shows that the International research maintains comparative equilibrium in four areas while Chinese research is not comparatively active on Applications.

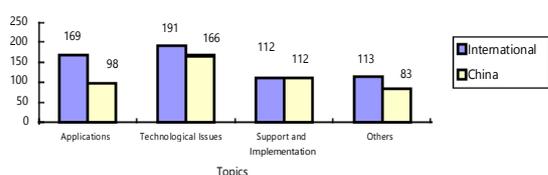


Figure 3. Distribution of articles by topic.

3.3 Distribution of Application Articles

Distribution of Application articles by year is shown in Figure 4. The international research was activated in 1996, and the Chinese research was intensified till the later half of year 1999. Since then, there appears no apparent difference in amount. It also shows that the International research increased rapidly at first, then stagnated for a few years because of the burst of the New Economy Bubble, and began to increase recently with e-commerce’s new development. Chinese research kept a fast increase trend with China’s e-commerce boom and the rapid rise of economy. Because CNKI had not updated the database during the time of our research, we are short of the data of November and December 2003, therefore it appears a sharp decrease in 2003.

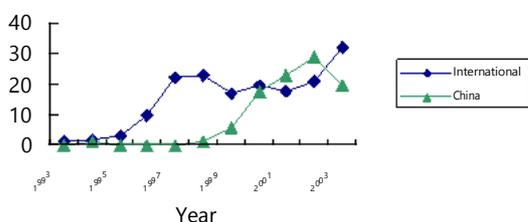


Figure 4. Distribution of Application articles by year.

From the figure of Distribution of applications articles, we can know that a large amount of this area is based on the IOS (International—52 articles, 30.8%; China—26 articles 26.5%), of which the majority is EDI systems. This is followed by topic on EC retailing

(International—30 articles, 17.8%; China—10 articles 10.2%), marketing and advertising (International—20 articles, 11.8%; China—22 articles 22.4%). Education and Training, Financial service and Auction also has increased greatly, such as the International research on Auction is 10.7% (18 articles). We can also know that International research has much more articles on IOS, retailing and auctions while Chinese research has the more on marketing and advertisements.

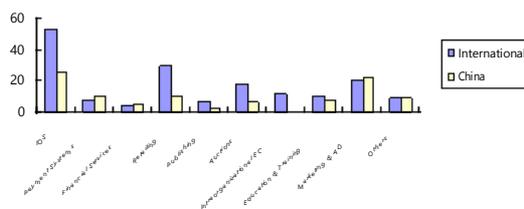


Figure 5. Distribution of Application articles.

3.4 Distribution of Technological Issues Articles

As shown in Figure 6, the research on Technological Issues is always increasing from the very beginning. This area has not been influenced by the Nasdaq Bubble generally, although the International research decreased in 2001 temporarily.

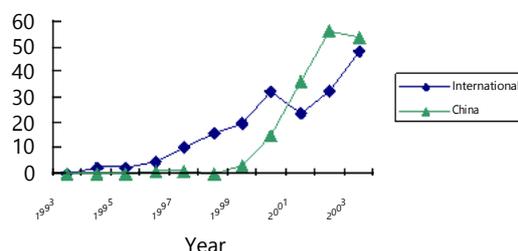


Figure 6. Distribution of technological issues articles by year.

From Figure 7, we can know that Support System is the very significant part for International (47 articles, 24.6%), and Chinese (46 articles, 27.7%) research. International research on technological component (54 articles, 28.3%) and Methodology/Algorithm (42 articles, 22%) is much more than Chinese researches. And Chinese research also has more articles on security (40 articles, 24.1%) than International researches. We can also know that Chinese research is very diverse in the area recently, and some researches are not limited to classification, therefore the research on others (39 articles, 23.5%) has a large percentage.

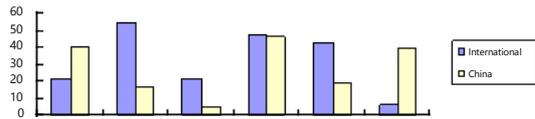


Figure 7. Distribution of technological issues articles.

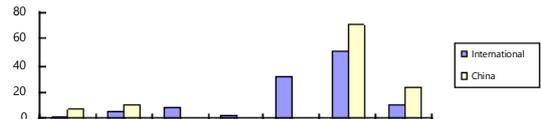


Figure 9. Distribution of support and implementation articles.

3.5 Distribution of Support and Implementation Articles

As shown in Figure 8, the research on Support and Implementation is always increasing with a small range from the beginning. After the burst of dot.com bubble, Chinese enterprises and researchers are trying to find the right corporate strategy and business model that are suitable to the conditions of China. As a result, support and implementation became a matter of concern and interest, and journals published heavily on this part. At the end of 2002, research on this part had been completely conducted to some extent. Advanced research is difficult to be conducted until the research on other part of e-commerce, such as new technology. Also because of the inexhaustive data of 2003, the curve reached its flattest level in 2000s.

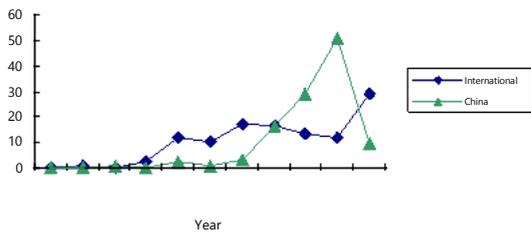


Figure 8. Distribution of support and implementation articles by year.

From Figure 9, we can know that Corporate Strategy is the most important part for International (51 articles, 45.5%), and Chinese (71 articles, 63.4%) research. E-commerce business strategy has to be re-created constantly in the ever-evolving world of cyberspace. For each enterprise, questions exist about how to develop the new business models, market strategy, what to do first, what to do next, and how to proceed. The answers are not always intuitive, and may not be reached the first time out. Therefore corporate strategy is always the focus of this part. International research on trust (31 articles, 27.7%) is skyrocketing recently, yet Chinese research still has no article on this area. Chinese research has much more research on taxation and legal issues than International research. On the other hand, it has no research on privacy, fraud, and trust.

3.6 Distribution of Others Articles

As shown in Figure 10, the research on others is increasing from the beginning. And the research became inactive between 1998 and 2000.

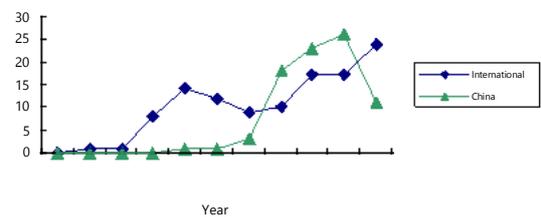


Figure 10. Distribution of others articles by year.

With the development of e-commerce, new concept, technology, and business model come up continually. International (73 articles, 64.6%) research on general introduction or about the foundation of e-commerce is more than general usage while Chinese research shows no apparent difference.

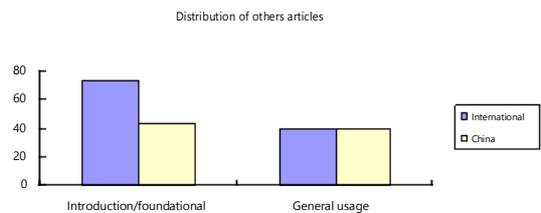


Figure 11. Distribution of others articles.

3.7 Distribution of International Articles

As shown in Figure 12, the International research increased from the beginning and Application was the most heavily published area. From 1999, research on Technological Issues ranked first and was followed by the research in the Application of e-commerce. The development and advance of technology made it impossible for e-commerce to appear and develop. E-commerce expanded its scope to more areas and showed the great impact to business and life, the research on applications had been heavily published during 1996 and 1999. Dot com and venture boom made e-commerce reach the peak and the progress of e-commerce also

required more technologies. Because of this business requirement, research on technological issues has become the focus of e-commerce research from 1999.

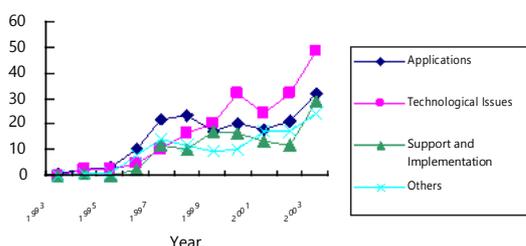


Figure 12. Distribution of International Articles

3.8 Distribution of Chinese Articles

From the figure of Distribution of Chinese articles by topic and year, we can know that although the research began later than International research, it advances more actively. The most distinguishing feature is that the research on Support and Implementation becomes more active than others. And the research on Technological Issues has ranked first from 2001 while the research on Application is inactive. The research on application is conducted comparatively less than International research; it might be the reason that the areas of application of e-commerce in China are not less widely. In order to activate the e-commerce development, it is necessary to conduct the researches on application much more widely and deeply.

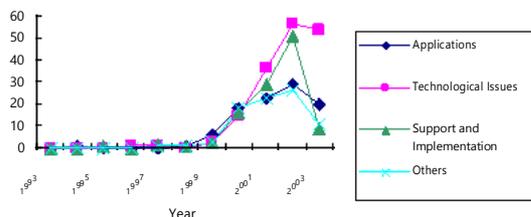


Figure 13. Distribution of Chinese articles

4. CONCLUSIONS

In our research, we reviewed fourteen International and Chinese journals and identified that 1,044 articles published between 1993 and 2003 are related to e-commerce. Although this review cannot claim to be exhaustive, it does provide reasonable insights into e-commerce. We can also get some important conclusions and implications through the comparative analysis of International and Chinese e-commerce research.

1. The e-commerce research will increase significantly in future although it might receive the influence of economic environment and other factors, because e-commerce will continue to grow at a fast pace and it

has become an important focus of public, researchers, governments, and international organizations etc.

2. Research has a close connection with the development of e-commerce. The business requirement and development of e-commerce can push the research; on the other hand, the research also can pull the e-commerce's development. The International and Chinese research and e-commerce have begun to activate with a remarkable interval. The International research and e-commerce activated from 1996, while Chinese research and e-commerce activated from later half of 1999.

3. In application area, the percentage of research on IOS keeps a decreasing trend while research on retailing, auction marketing and advertisement is increasing. This area will be more active and divers in the future with the development of online and mobile financial services, payment systems, and education.

4. Research on public policy needs much more concern. And new methods or standards should be developed for coping with problems in this area. There were not many public policy articles related to taxation, legal issue, privacy, and fraud. However, the research on trust has become a significant research area in e-commerce. And Chinese also started to conduct research on this area.

5. The differences between International and Chinese research can show the variness among e-commerce development phase, culture, and regional factors etc. to some extent.

6. China has less research on application, and it has already become obstacle for e-commerce. In order to activate the e-commerce, more researches are demanded to be done on these areas.

All reviewed articles that correspond to the classification scheme can be found at author's web site. <http://jaek.khu.ac.kr/bbs/articlebbs>.

REFERENCES

- [1] Alavi, M., P. Carlson, "A review of MIS research and disciplinary development", *Journal of Management Information Systems*, Vol. 8, No. 4, pp 45-62, 1992.
- [2] Kalakota, R., A.B. Winston, *Frontiers of the Electronic Commerce*, Addison-Wesley, MA, 1996.
- [3] Ngai, E.W.T., F.K.T. Wat, "A literature review and classification of electronic commerce research," *Information & Management*, Vol. 39, No. 5, pp415-429, 2002.
- [4] Nord, J.H., G.D. Nord, "MIS research: journal status and analysis", *Information & Management*, Vol. 29, No. 1, pp 29-42, 1995.
- [5] Turban, E., D. King, J. Lee, M. Warkentin, H.M. Chung, *Electronic Commerce 2002: A Managerial Perspective*, Prentice Hall, Upper Saddle River, NJ, 2002.
- [6] Walczak, S., "A Re-Evaluation of Information Systems Publication Forums", *Journal of Computer Information Systems*, Vol. 40, No. 1, pp 88-97, 1999.
- [7] Zwass, V., "Electronic commerce: structures and issues", *International Journal of Electronic Commerce*, Vol. 1, No. 1, pp 3-23, 1996.

A Conceptual Framework for E-tailing Quality and Its Evaluation

Jing Wang, Zhenquan Sha, Haiyuan Wang

School of Business Administration, South China University of Technology, Guangzhou 510640, China
wjchris@126.com

ABSTRACT

The Internet's influence in creating e-services has been revolutionary for retailer. Clearly, The Internet is fast becoming an important new channel for businesses in retailing, raising e-services as the emergent business paradigm in the industrialized world. Unfortunately, past research on measuring e-tailing quality is scarce. This paper provides a conceptual framework for e-tailing quality and discusses its evaluation. First, we summarize the main previous literature about service quality and e-service quality. Second, three analytical dimensions are identified: website quality, action quality and result quality, and each has its sub-dimensions. Each has different influence degree. Finally, we analyzing the evaluation way by comparing the online shoppers' perceived and expectation e-service. Future recommendations are also offered.

Keywords: E-tailing quality, E-service quality, E-service

1. INTRODUCTION

Since the 1960s, the Internet has been around, it was only a decade ago that the World Wide Web was born as a second segment of cyberspace. Undoubtedly, the digital revolution has changed almost every aspect of people's daily life, especially as stepping into the twenty-first century. In order to satisfy customers' sophisticated needs, many companies have set up their own web sites providing quality information and services to customers. When e-tailing was in its infancy, Web presence and low price were believed to be the drivers of success.

E-services, according to Roth [1], "are comprised of all interactive services that are delivered on the Internet using advanced telecommunications, information, and multimedia technologies." E-services are important in B2C e-commerce because they represent ways to provide on-demand solutions to customers strengthening the relations between customer and service provider, creating transactional efficiencies and improving customer satisfaction (Ruyter et al., 2001). While e-services are convenient and create efficiencies for their users, little is understood about how consumers evaluate them in operation and what are the important facts for customer choosing the E-services.

It is argued that better service quality can create better customer satisfaction (Zeithaml, Parasuraman and Malhotra, 2002). Consumers will satisfy or dissatisfy to a certain extent with the service, and will make the corresponding quality evaluation. This kind appraisal of service quality is actually a range degree function of quality between the consumer expectations and what they really get. However, according to an investigation [2], only 36% of the online shoppers are satisfied with their experience, and up to 28% of the online trade is end in failure, which prove the E-service Quality has already become one of the most important factors that

restrict customer to buy again. Nowadays, superior service quality has entered the picture as a differentiating strategy (Zeithaml, Parasuraman and Malhotra, 2002).

One of the main method for the website to be different from others, is to offer the higher service quality than its rival's. And the key is the service quality can satisfy the customer or exceed their expectations. In e-commerce, enterprises are unable to carry on the face-to-face communication with the customer, and almost all are offered through the website, here, it is impossible to carry on the proper adjustment by utilizing the feedback which through the oral communication with consumer even the physique language. In this case, it is very essential to understand the E-service Quality and how to evaluate.

2. LITERATURE REVIEW

Service quality has been recognized as an important strategic retailing weapon [3]. The most widely approach used to conceptualizing and measuring service quality has been developed by Parasuraman et al. (1985). Service quality has been defined as a form of attitude, related but not equivalent to the construct of satisfaction, which results from the customer's perception of service in relation to his or her expectations of service disconfirmation paradigm [4]. They also identify five dimensions of service quality and develop SERVQUAL, an instrument to measure customers' perceptions of Service quality [5]. The dimensionality and reliability of SERVQUAL has been the subject of many subsequent studies (e.g. Carman, 1990; Cronin and Taylor, 1992; Babakus and Boller, 1992; Cronin and Taylor, 1994; Parasuraman et al., 1994; Fischer and O'Bryan, 1995; Licata, Mowen and Chakraborty, 1995; Siu and Cheung, 1999, 2001; Siu and Woo, 1997 and 1999).

	Study	Focus	Dimension
↑	Berry, et al. (1988) and Parasuraman, et al. (1985, 1988)	Service quality	Tangibles; Reliability; Responsiveness; Communication; Credibility; Security; Competence; Courtesy; Understanding/knowing the customer; Accessibility
	Parasuraman, Zeithaml and Berry (1988, 1991)	SERVQUAL	Reliability; Responsiveness; Assurance; Empathy; Tangibles
	Harvey (1998) [7]	Service quality	Quality of actual results; Process to achieve those results (Consists of both technical and perceptual aspects)
↓	Kaynama, Black (2000) [8]	Modifying SERVQUAL in travel industry	Responsiveness; Content and purpose; Accessibility; Navigation; Design and presentation; Background; Personalization and customization
	Zeithaml, Parasuraman and Malhotra (2000, 2002)	E-SERVQUAL	Efficiency; Reliability; Fulfillment; Privacy; Responsiveness; Compensation; Contact
	Li et al. (2002) [9]	E-SERVQUAL	Responsiveness; Competence; Quality of information; Empathy; Web assistance; Call-back systems
	Yang, Peterson and Huang (2001) [10]	E-service quality	Ease of use; Content displayed on the Web site; Accuracy of content; Timeliness of response; Aesthetics; Privacy
	Lociacono, Watson and Goodhue (2000) [11]	WEBQUAL	Informational fit to task; Interaction; Trust; Response time; Design; Intuitiveness; Visual appeal; Innovativeness; Flow (emotional appeal); Integrated communication; Business processes; Substitutability
	Yoo & Donthu, (2001) [12]	SITEQUAL	Ease of use; Aesthetic design; Processing speed; Security
	Wolfenbarger and Gilly (2002) [13]	.comQ	Web site design; Reliability; Privacy/security; Customer service

Figure 1: Comparison of Previous Studies

Superior service quality is widely acknowledged as a driver of perceived value, which, in turn, will enhance customer loyalty (Parasuraman and Grewal, 2000) and improve the provider's image, sales and profitability (Buzzell and Gale, 1987; Gummesson, 1993). Therefore, the study on service quality has been focus on in many studies of various industries. While the e-business has

become a grown-up point of new economy, the research on Service quality have been combine with the e-commerce. Based on the previous studies, Zeithaml et al. (2000, 2002) developed the new concept E-service quality and used the e- SERVQUAL to measure it to study how customers judge e-service quality.

E-service quality is defined as seven dimensions: efficiency, reliability, fulfillment, privacy, responsiveness, compensation and contact, which form two scales: a core e-serve quality scale and a recovery scale [6]. The first four dimensions are classified as the core service scale, and the latter three dimensions are regarded as a recovery scale, since they are only salient when online customers have questions or problems.

Although the development of e-service quality notion and measurements is at an early stage, the studies on this field have developed fast and varied in the evaluation way. **Figure 1** has shown this concept developing process and summarized the main ideas.

3. DEVELOPING A CONCEPTUAL E-TAILING QUALITY FRAMEWORK

E-service quality has so wide range of all operation in e-commerce, including e-tailing, but the existing

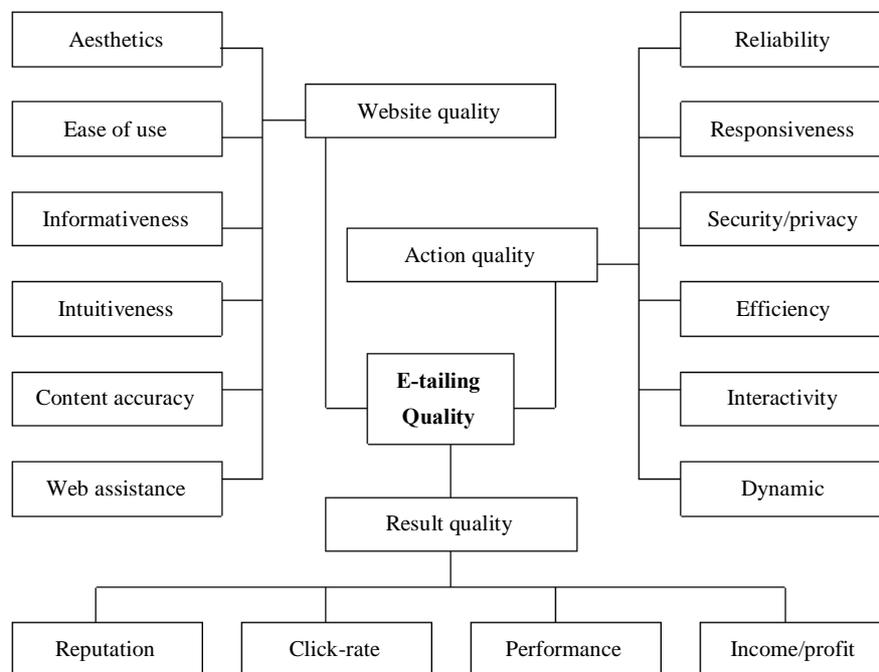


Figure2: E-tailing Quality Model

The quality of a web-based retail service depends on the perceived quality of the process of using the web for purchasing as well as the perceived quality of the outcome. E-tailing quality can be evaluated through three sections: website, action and result. Every section has serial dimensions. Website quality is mainly about the website design and customers' perception of the interactions that happen during their online-shopping. Action dimension deals with the process or the way in which service is delivered, and is main contributors of consumers' willingness to buy. That is, it reflects the outcome of the service or what is received. Result quality defines as the e-tailer image and quality. This integration may help identify the relationship between e-tailing quality and willingness to buy on-line.

literature is not yet rich enough to provide a conceptual foundation of e-tailing quality and because of that, it was required to develop an understanding of the detailed determinants in e-tailing quality and their impact on consumer attitude.

After extensive literature review, we develop the E-tailing quality framework to enrich the Web-oriented studies. Gronroos (1990) had proposed the concept of perceived service quality, and it depends the comparison between customers' expectation and the perceive service level in service quality. In addition, he divided service quality into " technical quality " and " functional quality " [14]. Here, we use his conceptualization of technical and functional quality for reference, combining the E-SERVQUAL and .comQ. All the literature led to the development of the conceptual model of e-tailing quality, which is displayed below as **figure 2**.

The website quality is a basic factor, which means is a go without saying function, so customer can get the request out of clearly, even not consider. If customer satisfies with it, it may not improve their satisfaction obviously. If the website can't satisfy them, customer may give up the online-shopping. Therefore, it is one of the importance competition factors. The action quality is an expecting factor, which the more realize in the online service, the more satisfied the customer is. When missing these demands, the customer is unsatisfied. Through offering it, the customer's perceived value of service can be improved, furthermore, improve the customer's loyalty. The result quality is a factor usually be neglected. The quality of the result e-tailer get can greatly affect the customer's perceived service quality. But all above isn't

absolute; it will change in different e-tailing product or service.

In the practical operation, it can evaluate the e-tailing quality according to measuring the consumer's expectation and perceived service. But the disparity between the actual and expecting e-service quality does not shape simply, instead, it finally forms by many kinds of disparity in the practices. Using the e-tailing quality model, reasons can be found when there are problems in e-service quality, and the proper method to narrow various kinds of disparity can be known, which offer the theoretical foundation for the e-tailer making the strategy and tactics of improving the consumer's perceived service quality.

The proposed model employs ETQ as the whole service quality in e-tailing, P_{ji} as the point of perceived service quality the i dimension get in the j section, E_{ji} as the point of expectation service quality the i dimension get in the j section, W_j as the weight of each section, W_i as the weight of each dimension, then as following:

$$ETQ = \sum_{j=1}^3 w_j \sum_{i=1}^n w_i (P_{ji} - E_{ji})$$

If more than one consumer give the evolutions, the average point can be used (N means the number of the sample):

$$AVETQ = \frac{\sum_{j=1}^3 w_j \sum_{i=1}^n w_i (P_{ji} - E_{ji})}{N}$$

What we need is the comparison between ETQ or AVETQ and 0, which is the comparison between the perceived e-service quality and the expectation.

4. CONCLUSION

In conclusion, this article proposes a e-tailing quality model based on the previous literature and the method to evaluate it. The objective of this article has been to present some preliminary frameworks for analyzing e-tailing quality, and the frameworks presented help clarify the discuss. The models of e-tailing quality is consist of three section: website quality, action quality and result quality. Each section has its dimensions. It provide ways to understand the affected factors in e-tailing service, and what is leading to the success. In order to construct the model, a careful review of previous studies has been operated.

However, there are numerous existing questions that require further inquiry in order to get best practices. The determined factors of e-tailing quality are changing under the e-business environment, and the customer's appraisal on e-tailing quality will be different as the kind and the characteristic of e-tailing are different either. How great degree can this three section and each

dimensions can reflect customer's expectation in ETQ, which factor is more important, whether there are other factors can affect ETQ, all are questions needed studying further.

REFERENCES

- [1] A.V. Roth, "Service strategy and the technological revolution", in: J.A.D. Machuca, T. Mandakovic (Eds.), *POM Facing the New Millennium: Evaluating the Past, Leading With the Present and Planning the Future of Operations*, Production and Operations Management Society and the University of Sevilla, Sevilla, pp. 159-168, 2000
- [2] Chang Guangsu, "Service quality in E-commerce", *World Standardization & Quality Management*, pp. 28-30, Jan, 2004
- [3] Chen, Z., and Dubinsky, A.J. "A Conceptual Model of Perceived Customer Value in E-Commerce: A Preliminary Investigation," *Psychology & Marketing* (20:4), pp. 323-347, 2003
- [4] Parasuraman, A., Zeithaml, A.V., Berry, L.L., "A conceptual model of service quality and its implications for future research", *J. Marketing* 49, 41-50, 1985
- [5] Parasuraman, A., Zeithaml, V.A., and Berry, L.L. "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing* (64:1), pp. 12-40, 1988
- [6] Zeithaml, V.A., Parasuraman, A., and Malhotra, A. "A Conceptual Framework for Understanding E-Service Quality: Implications for Future Research and Managerial Practice," *Working Paper No. 00-115*, Marketing Science Institute, Cambridge, MA, 2000.
- [7] Harvey, J., "Service quality: a tutorial", *Journal of Operations Management* 16, 583-597, 1998.
- [8] Kaynama, S.A., and Black, C.I., "A proposal to Assess the Service Quality of Online Travel Agencies: An Exploratory Study," *Journal of Professional Services Marketing* (21:1), pp. 63-88, 2000
- [9] Li, Y.N., Tan, K.C., and Xie, M., "Measuring Web-based Service Quality," *Total Quality Management* (13:5), pp. 685-700, 2002
- [10] Yang, Z., Peterson, R.T., & Huang, L., "Taking the Pulse of Internet Pharmacies", *Marketing Health Services*, Summer, 5-10, 2001
- [11] Loiacono, E.T., Watson, R.T., Goodhue, D.L., "WebQual: A Website Quality Instrument," *Working Paper 2000-126-0*, University of Georgia, 2000.
- [12] Yoo, B., and Donthu, N., "Developing a Scale to Measure the Perceived Quality of an Internet Shopping Site (SITEQUAL)," *Quarterly Journal of Electronic Commerce* (2:1), pp. 31-45, 2001
- [13] Wolfinbarger, M.F., and Gilly, M.C., ".comQ: Dimensionalizing, Measuring and Predicting Quality of the E-tail Experience," Working Paper No. 02-100, Marketing Science Institute, Cambridge, MA, 2002.
- [14] Gronroos, C., "Service Management and Marketing: Managing the Moments of Truth in Service Competition", *Lexington Books*, Lexington, MA, 1990.

A Multi-Agent System for E-Business Processes Monitoring in a Web-Based Environment

Xinyu Zhao, Cen Wu, Runjie Zhang, Chen Zhao, Zuoquan Lin

Department of Information Science, Peking University, Beijing 100871, China
{xinyua, wucen, zrzj, zchen, lz}@is.pku.edu.cn

ABSTRACT

In this paper, we present a multi-agent system MAGS for the e-business processes monitoring in a web-based environment. We classify the types of agents in MAGS by their monitoring capabilities. An algorithm is given to explain the mechanism of supervising and controlling the execution of business processes. An abstract model of alerts, which can give warnings of infringement on business policies, is proposed. Access control can also be realized by MAGS, which manifests in delivering different view of the business process to different roles participate in it. Being successfully adopted in a customer service management system, MAGS has been proven flexible and practical.

Keywords: Business Processes Monitoring, BDI, Agent, Capability, Web Services

1. INTRODUCTION

A trend of e-Business is to wrap legacy applications with web services technologies, by allowing enterprises to integrate internal existing processes with the trading partners' in a manageable manner. As such, a business process specifies a potential execution order of operations from a collection of web services involving multiple organizations. With the number of complex business processes increasing, automated monitoring technology is in great demand. However, the traditional passive, centralized business process management system is not applicable for fully distributed applications in a web-based environment. For example, without initiative supporting and monitoring, it is hard to response timely and accurately to numerous events from other applications inside and outside the enterprise. Moreover, the detailed information about distributed business processes should be gathered by the requirements of different users and the relationships with the particular business process instances. Nowadays most traditional systems can not deal with the above issues well.

Software agent is an object which has autonomous actions of accomplishing specific task [1]. In multi-agent system, an agent interacts with other agents or environments to achieve its goal by communication, coordination and cooperation. Of various agent architectures which have been presented, BDI model [2] [3] [4], is probably the most mature and has been adopted in a number of research and industrial applications. There are some works introducing multi-agent into the execution monitoring of business processes [5] [6] [7]. The rationale-based monitors [5] aim at planning and removing the conventional assumptions in static and determinate. Classifications of monitors are proposed to suggest plan transformations. Execution Assistants [6] is implemented to assist human in monitoring robots' behaviors. A top-level categorization of alerts is presented, but the next lower

level of the ontology is not provided. Execution Assistants emphasize the VOI (Value Of Information) and VOA (Value Of Alerts) and analyses the relationships between the two domains. Continuous Planning and Execution Framework [7] defines a monitor to be an event-response rule. Taxonomy of monitor classes derived from the types of events is believed to enable simpler and more modular specifications of monitors. Considering in the new web environments, in which multiple services and applications interact with each other, richer monitoring capabilities, such as alert technology and access control, are in demand for the monitoring management of the business processes.

This paper presents a multi-agent system MAGS (*Monitoring AGent System*) for the monitoring management of the business processes in a web-based environment. Conventional BDI agent model is extended to BDIC (Belief, Desire, Intention and Capability) model with monitoring capabilities, which are induced from a real application. MAGS is assigned to supervise and control the business processes following the profile of requirements of administrators. Also, MAGS is able to continually monitor the incoming system events, and then carry out alerts to the corresponding service providers accurately and timely via multiple channels. MAGS is able to cooperate with other agents to get the role of the users involved in the business process instance and give the detailed information about the instance. We will demonstrate how the implemented MAGS is integrated in a customer service management (CSM) system and show some of its novel features.

This paper is organized as follows. In Section 2, an example drawn from a distributed call center is depicted to show our motivation throughout this paper. Section 3 presents the architecture of MAGS and the BDIC agent model. In Section 4 the implementations of MAGS are discussed. Finally, conclusions are drawn and future works are discussed in the concluding section.

2. MOTIVATING EXAMPLE

We will illustrate our motivations with a distributed call center example. Internet-Enabled Customer Services [8] have become the most important channel, through which customers can talk directly to the corporations. Customers can call and state their requests to the distribution corporation. When the call center receives a call from a customer, the broker of call center divide the requirements into different service types, e.g., advisement, repair and so on. The broker should find a business partner, for example Service Net, from which customers can get the services, to fulfill the customer's request. When the proper business partner is found, the broker will make an appointment with the customer. Meanwhile, the customer's information and requirement are dispatched to the Service Net via multiple channels, such as short message service (SMS), email, web and fax, etc. After the requested service is finished, the Service Net submits a service report to the call center. To keep a satisfactory service level, the call center will collect the customers' feedbacks and track the Service Net's performance measure daily. If there are customers making complaints, their requirements will be re-accepted. At the same time, the department of service quality control will evaluate the quality of the whole business process. A typical business process of dealing with the customer incoming call is shown in Figure 1. Due to the distributed properties of the system, the business processes might not be executed as we expected. Without automated support, the call center cannot deal

with numerous unexpected events correctly and timely. Usually, it might be required that 95% of received calls should be answered within 10 seconds. Automated alert technology is required to give the warning of exceptions. Different roles participating in the business processes, such as call center, Service Net and service quality control department, have different information requirements. The details of running and completed business processes should be given based on the roles of the requester. All above motivate us to make an investigation on the monitoring technology for cross-organizational business process in web environments.

3. MAGS ARCHITECTURE

3.1 BDIC Monitoring Agent Model

A belief-desire-intention-capability (BDIC) architecture includes an explicit representation for an agent's beliefs, desires, intentions and capabilities. In this paper, it's the capability component that determines agent's types and monitoring functions. We concentrate on monitoring mechanisms here and the details of agent capabilities can be found in [9]. We give the definition of BDIC Monitoring Agent first.

Definition 1 BDIC Monitoring Agent is a structural Monitoring Agent = (B, D, I, C, Se, Ef, Pr), where (see Figure 2.):

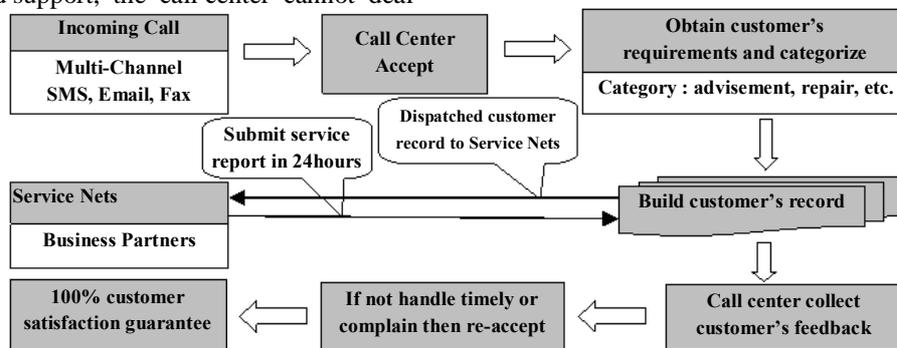


Figure 1 Closed loop diagram of an incoming call business process in call center

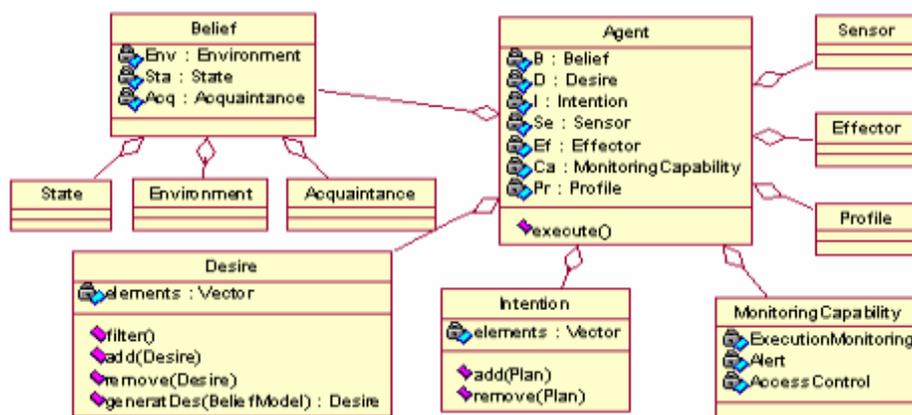


Figure 2 UML diagram of monitoring agent

B is the beliefs of monitoring agent, that is, the informational state of the business process in which the agent is monitoring. The beliefs include State, Environment and Acquaintance.

D is its motivational state, that is, what the monitoring agent is trying to achieve.

I is the intentions of monitoring agent, that is, the planning for eventual execution.

C is monitoring capability including categories of Execution Monitoring, Alert and Access Control. It will be further explained in the next sections (see Figure 3).

Se is the sensors of monitoring agent. It sensors information from environments or other agents by the standard message format e.g. FIPA, ACL.

Ef is the effectors of the monitoring agent by which the monitoring agent pursues its intentions.

Pr is the information requirements profile of system users.

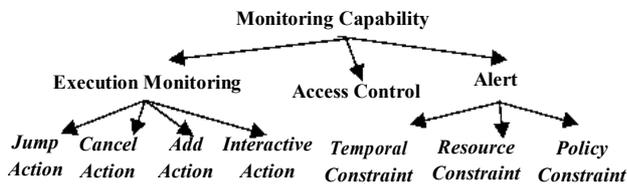


Figure 3 Monitoring capabilities

Monitoring agent is a typical BDI agent whose sensor perceives the information of business processes and revises beliefs. Desires is formed based on the current mental states (*B*, *D*, *I*). Profiles are customized by the system users. Monitoring agent pursues intentions by effectors. The type of agent is determined by its monitoring capabilities.

following definition.

Definition 2 A business process is a structural BP = {precondition, effect, body}.

3.3 Execution Monitoring Agent

Below is an agent monitoring-interpreter, which explains the mechanism of the execution monitoring of business processes.

```

Monitoring-Interpreter
Initialize-state();
repeat
  if all of currentBP.effects are true
  then currentBP = getNextBP() // Jump Action
  end if
  if one of currentBP.precondition is false
  then tempBPList = getBPList()
  // Get a list of business processes that will make the
  // false preconditions of the current business process true.
  if tempBPList == null
  // Cancel Action: no business process will
  // make the false preconditions of the current business process
  // true.
  then reportToAdmin()
  currentBP = getNextBP ()
  else currentBP = getBestBP(tempBPList)
  //Add Action and Interactive
  Action: interact with human to get the best business process
  that can achieve the goal in high quality.
  end if
end if
execute(currentBP.body)
end repeat
    
```

According to the monitoring-interpreter, the execution monitoring agent has the following actions:

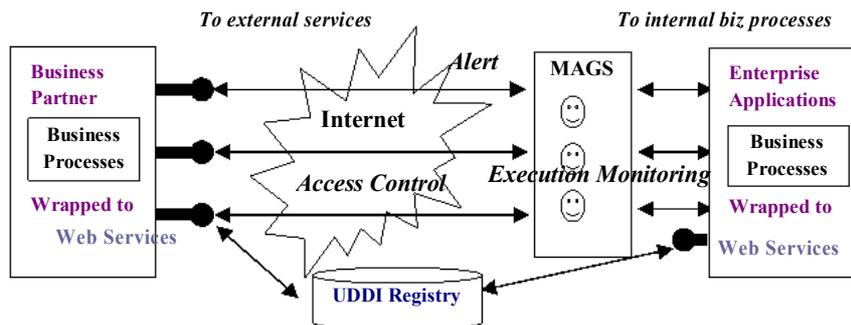


Figure 4 Integrated architectures

3.2 Integrated Architectures

In order to make the following explanations explicit, we give a kind of definition for the business processes here. A business process is a set of one or more linked procedures that collectively realize a business objective. By packaging business processes as services that are accessible over the internet, enterprises achieve new and better means to utilize their own and other's applications (see Figure 4). In general, a business process includes preconditions, effects and execution body as the

Jump Action: If all effects of current business process are true in the dynamic environments, it is not necessary to execute current business process in normal sequence. Whenever the effects of current business process are already true at the beginning of the execution, the execution monitoring agent will perform a jump action which changes the business process engine's focus of attention to the next process.

Considering the case in our application, when the broker has accepted the calls, yet not dispatched it to the Service

Net, the customer calls again to tell that he has solved his problem. It means that the goal of the Dispatch process has been achieved, thus the execution monitoring agent leads the Feed-Back process to be the current process.

Add Action: Changes of the environments may make the current business process preconditions false so that it could not be executed. The execution monitoring agent watches the preconditions continually. If one of preconditions fails, the execution monitoring agent will add a new process to execute which can make the false precondition true. Thus current business process can keep going on.

Cancel Action: During the execution of current business process, its preconditions may become infeasible due to the unpredictability of environments. Going on business processes with infeasible preconditions, the system may throw unwanted exceptions. The execution monitoring agent will watch the preconditions of current business process. If one of precondition cannot be made true through any Add-Actions, the execution monitoring agent will report to administrators and cancel current business process.

Interactive Action: In real-life domains, there may be multiple processes available for achieving a certain goal. It requires human experience and judgment for the business processes to execute in a high quality. Whenever multiple choices appear, the execution monitoring agent will interact with system users to get their knowledge and choose the most suitable business process to execute.

3.4 Alert Agent

A business process is usually constrained by policies or rules that must not be violated during its execution. Alert agents give warnings of infringing on business policies and try to avoid the failure of business process.

Definition 3 Alert is a structure $Al = (Id, Re, Ty, Le, Ch, Co)$, where:

Id is the unique identifier of the alert.

Re is the receiver of alerts, which commonly are services provided by business partners.

Ty is the type of alerts, including temporal, resource and policy constrains.

Le is the emergent level of alerts.

Ch is the channel of alert sent through.

Co is the content of alert.

When the business policy is infringed during the business process execution, the alert agent sends an alert to the specific service provider. An alert has not only the identification information but also an emergent level of indicating its urgency. MAGS support alert agents to send alerts through different channels. Alerts are classified into three types by temporal, resource and

policy constrains.

Temporal Constraints: The business processes are commonly demanded to execute under temporal constraints. Each temporal constraint consists of a sequence of actions and one or more pairs of temporal constraint ontology and time. The temporal constraint ontology and its meanings are shown in Table 1. Temporal constraint has the following form:

BPName TC_Ontology Time [&TC_Ontology Time...] action1 [; action2...]

It indicates that the business process with the name of *BPName* must not violate the temporal constraints; otherwise the alert agent will perform corresponding actions. The alert will also be composed and sent by alert agent.

Table 1. Temporal ontology

TC_Ontology	Meaning
Start-at	Start time must be at given time
Start-lt	Start time must be later than given time
Start-et	Start time must be earlier than given time
End-at	End time must be at given time
End-lt	End time must be earlier than given time
End-et	End time must be earlier than given time

In our domain, CSM provides interfaces of getting the entire view of the business processes for system users. Generating the entire view of a business process is a complex and time-consuming process and it is restricted by time constraints. The following is a temporal constraint in the profiles of alert agent:

GetBPView Start-et 8:00:00&Start-lt 20:00:00 reportToAdmin; rejectRequest

According to the constraint, the business process *GetBPView* should start earlier than eight and later than twenty everyday. It makes the *GetBPView* process stagger the rush hour with other business processes and ensure CSM safety. If a system user attempts to invoke the *GetBPView* process in the interval of 8:00:00 and 20:00:00, the alert agent will compose an alert to administrators and reject the request.

Resource Constraints: The business processes in a resource-bounded environment should not overspend the system's resources. Resource constraint has the following form:

BPName resource threshold [&resource threshold...] action1 [; action2...]

It indicates that resources spent by the *BPName* process must not exceed by the threshold; otherwise the alert agent will perform corresponding actions to give an alert.

Policy Constraints: The business processes must maintain the business policies that must not be violated. Alert agent gets policies from its profile and continually monitors the incoming events unsatisfying conditions in the policy constraints. If any constraint is violated, the alert agent will perform corresponding actions to give alert. Policy constraint has the following form:

BPName condition [&condition...] action1 [; action2...]

(RBAC) system and can cooperate with MAGS in offering the desired security. The access control agents communicate with WebDaemon through Authentication and Authorization Protocol (AAP). Based on the user's role returned by WebDaemon, the access control agent can give the corresponding view of running or completed business processes to the particular user.

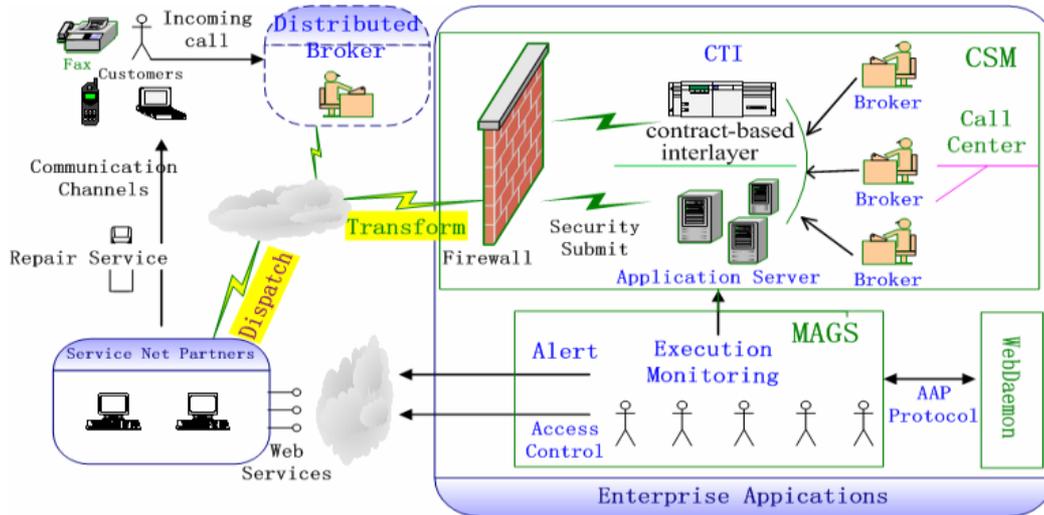


Figure 5 Overall implementations of enterprise applications

Operation	Emergent Degree	BPID	Name	VIP	AcceptDate	Phone	Product	RequireDate	RequireTime	Type	ServiceNet	BPSState
Details	Level 3	200309123	M. Jack	Y	2003-09-01 09:47:40	62761817	TV	2003-09-03	Before 17:00	Repair	BJ Car	Accepted/Dispatched
Details	Level 1	200309232	D. Jones	N	2003-09-02 11:27:34	62361835	TV	2003-09-04	Before 12:00	Repair	Xitian	Accepted/Dispatched
Details	Level 2	200309446	Wang Zhengqin	Y	2003-09-02 14:53:45	86753416	Fridge	2003-09-03	Before 17:30	Repair	BJ Car	Accepted/Dispatched
Details	Level 2	200309532	Chen Yang	Y	2003-09-02 15:53:43	82763416	Washer	2003-09-03	Before 17:30	Repair	BJ Car	Accepted/Dispatched

Figure 6 Web pages of detailed Dispatch process information reported by MAGS

3.5 Access Control Agent

From the integrated MAGS architecture (Figure 4), we know that there are different kinds of requirements of business process information from inside the enterprise, e.g., supervisor, and from outside the enterprise, e.g. the business partners. There are problems if everyone can access no different information from the same process. In order to solve the security problem, the access control agents in MAGS expose interfaces to provide this function. WebDaemon [10] is a role-based access control

4. IMPLEMENTATION

The multi-agent system presented here has been successfully implemented and seamlessly integrated in a Customer Service Management (CSM) system of a company¹ (see Figure 5.). Agents act as intelligent session Enterprise Java Beans (EJB) within MAGS. Under the J2EE framework, session beans are components containing business logics associated with a

¹ TCL Group Corporation, <http://www.tcl.com>.

particular client session or task. MAGS runs on a J2EE application server, namely Weblogic, and adds intelligence in the sense of intention-based monitoring behaviors. CSM adopts a correlative work, called the contract-based interlayer [11], to integrate the layer of hardware-related functions (SMS, FAX) with the layer of business-related functions. With the contract-based interlayer, MAGS can send out alerts to business partners' services. MAGS cooperates with WebDaemon in providing access control functions. The statuses of the whole business processes can be monitored by web browsers based on the J2EE framework (as Figure 6 illustrates). We have got reliable and effective results in practice.

MAGS targets at providing a mechanism for business process management in a web services environment. MAGS is loosely coupled with other applications in architecture in the sense that it can be plugged into any potential systems, including e-commerce, web services and web-based applications. There is no implementary obstacle for these systems if only they hold out the specification of web services.

5. CONCLUSIONS

In this paper, MAGS, a monitoring multi-agent system for the management of business process in a web services environment, has been presented. MAGS can be integrated with applications in the enterprise and web services provided by cross-organizational business partners in architecture. Capabilities are added to BDI agent in MAGS to offer mechanisms of monitoring process execution, carrying out alerts and access control, which is described in detail with the examples of a typical business process in a customer service management system.

In the future, we will focus on developing richer monitoring capabilities and introducing MAGS into more domains, such as Semantic Web Services [12]. Semantic Web Services, in essence, is a kind of business processes. Its Semantic Markup OWL-S [13] should provide declarative descriptors for the state of execution of services. In fact, versions of OWL-S developed so far have not ventured into this area. In our view, introducing semantics into MAGS is a promising approach to make functionalities of automatic monitoring possible.

ACKNOWLEDGEMENT

This work is supported in part by the National Science Foundation of China under the contract numbers

60373002 and 60496323, and in part by TCL group corporation. The authors thank for all the members of the CSM project.

REFERENCES

- [1] Woodridge M, Jennings N.R., "Intelligent Agent: Theory and Practice", The Knowledge Engineering Review, Vol.10, No.2, pp115-152, 1995.
- [2] Michael Bratman, *Intentions, Plans, and Practical Reason*, Cambridge: Harvard University Press, 1987.
- [3] P.R.Cohen, H.J.Levesque. "Intention is choice with commitment". *Artificial Intelligence*, Vol. 42, No.2-3, pp213~261,1990.
- [4] Rao, A.S. & Georgeff, M.P. "BDI Agents: From Theory to Practice". proceeding of the First International Conference on Multi-Agent Systems (ICMAS'95), 312-319.
- [5] Veloso, M., and Cox, M., "Rationale-based monitoring for planning in dynamic environments". Proceeding of the 1998 International Conference on AI Planning Systems, pp171-180. AAAI Press, Menlo Park, CA.
- [6] David E. Wilkins, Thomas J. Lee and Pauline Berry, "Interactive Execution Monitoring of Agent Teams", *Journal of Artificial Intelligence* Vol.18, pp217-261, 2003.
- [7] Kaminkaa, G., Pynadath, D., & Tambe, M. "Monitoring deployed agent teams". proceeding of *Autonomous Agents'01*, pp. 308-315, Montreal, Canada.
- [8] The Gartner CRM Vendor Guide: 2004, <http://www4.gartner.com> 5 April 2004
- [9] Xinyu Zhao, Zuoquan Lin, "An Agent-based Supply Chain Modeling", *Computer Science(In Chinese)* Vol. 8, 2004
- [10] Yanfeng Gui, Zuoquan Lin, "WebDaemon: A Role-based WEB Security Access Control". *System Computer Research and Development(In Chinese)*, Vol.8 pp20-30, 2003
- [11] Wu Cen, Zhao Xinyu, Zhao Chen, Lin Zuoquan, "Contract-based Interlayer: A Two-way Approach to Integrate Call Center With J2EE Framework", *International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES 2004)*.
- [12] S.A. McIlraith, T.C. Son, and H. Zeng, "Semantic Web Services," *IEEE Intelligent Systems*, Vol. 16, No. 2, pp46-53, Mar./Apr. 2001
- [13] DARPA Agent Markup Language. DAML-S/OWL-S 1.0 Draft Release. 2003. [Http://www.daml.org/services/owl-s/1.0/](http://www.daml.org/services/owl-s/1.0/)

A Study of Competences and Indicators for Electronic Commerce Professional Managers

Chihhung Wu¹, Wenchang Fang², Gaoliang Wang¹

¹ Department of Business Administration, Takming College, Taipei, 11451, Taiwan, China

² Department of Business Administration, National Taipei University, Taipei, 110, Taiwan, China
ericwu@takming.edu.tw, fang@ntpu.edu.tw, wanggl@takming.edu.tw

ABSTRACT

The objective of this research is to investigate the managerial competences for Electronic Commerce professional managers. Three research approaches have been adopted here: Focus Group, Fuzzy Delphi Method. The results indicated that professional skills and leadership competence are more important than other competences such as administrative skills and motivation. Comparing the competences of traditional managers and EC managers shows that the main difference is the emphasis on EC managers' professional skills.

Keywords: Competence of Electronic Commerce, Fuzzy Delphi Method

1. INTRODUCTION

While there have been quite a lot of studies on electronic business related issues, there are still very few researches on the competences of Electronic Commerce professional managers. The earliest research on competence was in late 1960s, which focused on the study of personality. Recent studies on competence, however, have changed the direction from the conventional understanding of competence as the quality, knowledge and skills one possesses to the notion that competence means not only the qualities one possesses but also those one does not own and therefore needs to learn. Murakami (1988) suggested that competence lies in the concrete performance one provides in the profession, which reflects on his/her substantial ability of achievement. Spencer & Spencer (1993) affirmed that competence is the "underlying characteristic" one has. They went further that the underlying characteristic is not only related to one's profession and position at work, but predictably or even actually affects his/her behaviors and performance. Yeung (1996) also indicated that the critical key to improve one's working performance is competence, which means the specific basic quality in a person. In addition, Ritter and Gemunden (2004) stated that both network competence and technological competence have a significant positive impact on a company's innovation success. Mansfield (1996) added that competence is the description of accuracy, skillfulness, and specific behaviors, which needs to be studied by the staff, so that they will have good performances and better achievements at work. Besides, as pointed out by Tseng (1998), competence is the crucial factor of achievements. Therefore, in spite of the fact that scholars from each school have had different definitions on competence, most scholars tended to agree that competence is closely related to one's profession, which is also the crucial key to one's performance and achievement at work.

The contribution of our study is to provide an insight into the competences of EC professional managers. It will offer the companies and industries a clear direction for training and hiring managers, which will also promote the managers' outstanding achievements. Besides, the indicators for EC professional managers can be adopted as the indicators for enterprises to evaluate the competences of their managers. Accordingly, this study proposed to achieve the following goals: 1. To examine the constitution of professional competence possessed by successful EC professional managers. 2. To find the indicators of competences for EC professional managers.

2. LITERATURE REVIEW

This paper discusses the aspects and indicators of managerial competences for EC professional managers. The first part of this literature review presents studies related to managerial competence, while the second part focuses on the methods of evaluation for competence.

2.1 Studies on Managerial Competence

Although the earliest study on competence appeared in late 1960s, recent studies on competence rarely focus on the research of managerial competence possessed by EC managers, and without mentioning the indicators for them. This study aims to start with the previous studies on managerial competence and managerial skills both in Taiwan and overseas. In this case, the managerial competence which needs to be possessed by EC managers be investigated, so that the research methods and procedure could be made. To this purpose, contemporary studies on managerial competence will be discussed as follows: (1) the personality trait school, (2) the functional school, (3) the situational school, and (4) the survey school.

The personality trait school tended to adopt "the moment of heroism" and suggested that successful

managers should have some special characteristics in their personalities. Hence, as long as a person has special characteristic in his/her personality, he/she could be trained and become a successful manager. In terms of personality, Gattell (1973) has made observations and analyses of various behaviors in our daily life. Eventually, the 171 types of personalities he found were sorted into 16 categories according to factors that would affect people's behaviors. Robbins (1985) tended to see from the managerial perspective and asserted that a person's ability is mainly composed of intellectual and physical abilities. Intellectual ability means the ability of mind activities, which includes the abilities of calculation, language, response, and logic. Physical ability denotes one's spirit, physical flexibility, energy etc. Professions of higher positions require more intellectual ability, whereas technical or mechanical jobs need people with more physical ability. Maslow (1954) proposed the hierarchy of needs theory and suggested that there are five levels of needs for every one. From the lowest-order need to the highest-order need they are: physiological need, need of safety, social need, need of self-esteem, and need of self-actualization. In general, for those who require less in life, physical need tends to be more important, while for managers the needs of self-esteem and self-realization would be more emphasized. However, there are quite a few uncertainties in the personality theory, such as the neglect of interactions between leader and staff, the fact that personalities are too complicated and difficult to be clarified, etc. Hence, theories in this school have gradually tended to be ignored.

Katz, the leader of the functional school, argued that managers of various levels require different levels of skills such as following: 1. professional skill, 2. social skill, and 3. technical skill (Katz, 1955). Those three skills would be different due to the different tasks required for managers of different levels, which was also confirmed in the theory of Guglielminio (1979). In other hand, the situational school emphasized the importance of circumstances, and believed that some specific situations could make some people become heroes. In other words, they denied the importance of the manager's personality and function. They took further that what makes one a successful manager was not necessarily his/her personality, but how to deal with the confronted situations. For instance, Fieldler (1967), a scholar of this school, proposed the leadership contingency model and suggested that the function of the leadership could be influenced by the relationship between the leader and the fellow group members, the structure of each task, and the authority of the leader. Besides, in his path-goal theory, House (1971) added that the manager's behaviors can be affected by the circumstantial factors, such as the structure of the work, the system of the official authority, and team involved in the work. Leadership can also be affected by the quality of staff members, such as their ability for internal and external control, experiences, and abilities.

All these factors would have influences on the manager's performance and satisfaction with the job.

The proper managerial competences for the managers were discovered via survey methods in the theory of investigation. Chang (1998) surveyed managers at various levels and suggested that the important managerial competence for managers at higher positional levels are "planning and coordination", "decision making", "management reforming", "strategies adopting", "prospect", "problem solving", "sensitivity to the market", "consultation and authority", "insurance managing", "social network", "outstanding performance", and "perfectionism". The important managerial competences for managers of intermediate positional levels are: "project and procedure management", "goals setting", "communication skills", "staff training and supporting", "incident managing and group constructing", "innovation" and "negotiation". The major managerial competences for managers of basic positional levels are: "professional skills", "efficiency", "being active and enthusiastic", "executive ability", "time management", "quality management", "reliability", "good learning attitudes", "customer services" and "emotional control". Chung (1998) has researched on the managerial competence of higher positional managers in the high-tech industry, and his research focused on 159 junior managers of higher positional level in an industrial technology research institute (ITRI). The result showed that competences possessed by managers there were the abilities of "strategies adopting and coordinating", "problem solving", "quality control", "action directing", and "group constructing". The major competences that needed to be trained and improved are "sensitivity in management", "ability of negotiation", "strategies adopting and coordinating", "inspiration on others", "quality control". Wu (1993) asserted that managerial competence should include both personal qualities and managerial skills. In terms of personal qualities, personality and motivation would be the main factors, whereas managerial skills contain: social skills (including communication skills), administrative skills (i.e. planning and coordinating), cognitive skills (such as analysis and decision making), leadership skills (such as inspiring and tutoring), and technical skills.

Recent studies on managerial competence have tended to include theories and perspectives from different theories and have combined various methods to examine managerial competences. Ritter and Germunden (2004) discussed and analysis the impact of the dual nature of the key to competitiveness in the network economy. Lin (1996) has adopted the Fuzzy theory to the evaluation models of managerial competences, considering aspects on both personal qualities and skills. Personal qualities include ability, personality, and motivation, whereas managerial skills could be divided into the aspects of technical skills, social skills, and conceptual skills. Additionally, Wu

(1999) has also adopted the Fuzzy Analysis Hierarchy Process to construct the evaluation model of managerial competence and suggested that managerial competence contains eight aspects: (1) conceptual skills, (2) social skills, (3) leadership skills, (4) administrative skills, (5) professional skills, (6) personal qualities, (7) abilities and (8) motivations. Accordingly, the direction of his research has already covered the concepts from personal characteristic, functional and investigative theories. Chang, Huang, and Lin (2000) proposed a new fuzzy Delphi method and fuzzy statistics for managerial talent assessment. Their taxonomical approach was derived from the open systems view of firms. The new method was able to identifying and distinguishing different aspects of organizational competence. Sanchez (2004) also developed a taxonomy of five models of competences than an organization must develop and maintain in its various activities to reach overall competence.

2.2 Methods of Evaluation on Competences

The primary methods of evaluation on competences are as follows: 1. Evaluation on working ability (McClellan, 1973; Tseng, 1998). 2. Focus group. 3. Examination on competence models (Hsieh, 1998). 4. Peer Evaluation (Parry, 1998). 5. Assessment centers (Parry, 1998). 6. Circumstantial tests (Parry, 1998; Chang, 1997). 7. Evaluation tables of competence (Chuang, 1997). 8. Fuzzy theory (Ishikawa, et al., 1993; Lin, 1998; Wu, 1999). Apart from these, Richard & James (1995) had adopted the two-stages method which combined qualitative and quantitative methods for the research of competencies for computer software technicians. While in the first stage qualitative study approach was adopted for detailed interviews, in the second stage quantitative analysis was used to generate the final list of competence via Q-Sort techniques and discriminant analysis.

Ishikawa et al (1993) had adopted the Fuzzy theory for the purpose of solving some problems in the existing conventional Fuzzy Delphi Method. They used Max-Min and Fuzzy Integration to examine Fuzzy Delphi Method's workability, and the result has proven that the Fuzzy method has some advantages that the traditional methods do not have. Therefore, more and more scholars, such as Lin (1996), have started to adopt Fuzzy theories in the evaluation of managerial competences. Later, Wu (1999) and Huang, Chang, and Lin (1997) constructed his evaluation model of managerial competences through Fuzzy Analysis Hierarchy Process.

To sum up, from the literature review we found that in terms of competence there were no exact common consensus between theories from different schools, and the methods adopted were various. Therefore, it would be important to start with a more complete managerial structure, which should also be adoptable for electronic

commerce. Since the managerial structure proposed by Wu (1999) has included more aspects, our study proposes to adopt his structure of competences as the initial structure for the study of managerial competences, which is illustrated bellow. In terms of methodology, to increase the accuracy of our research and to follow the tendency of studies nowadays, we have adopted several methods and the combination of qualitative and quantitative methods. The Fuzzy Delphi Method used to choose suitable aspects and indicators of the managerial competences for EC managers.

3. METHODOLOGY

The objective of this research is to discover the managerial competences and indicators for EC managers. The initial construct of competences for EC professional managers is shown in Fig. 1. To achieve this goal, a both qualitative and quantitative approach is adopted in our research. In this way the advantages from each method could be integrated in this study, instead of relying too much on a certain type of methods and the disadvantages that might occur. Here, the process of this study could be discussed in two parts: 1. focus group, and 2. Fuzzy Delphi Method.

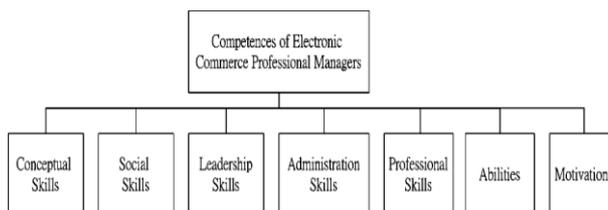


Figure 1. The Initial Construct of Competences for EC Professional Managers

3.1 Focus Group

After establishing the structure through literature review, a focus group of EC experts was held on for the purpose of discovering the managerial competences related to the job of EC managers, which also provided some suggestions for the adjustment to the initial structure of our study. To ensure that the experts invited would be representative, 12 experts from the industrial, governmental, and academic fields of the electronic commerce in Taiwan were invited for the focus group. The experts from the industrial field were the leaders who were responsible for electronic commerce business in a certain enterprises.

3.2 Fuzzy Delphi Method

From the opinions collected via the focus group for the adjustments on the items and aspects of managerial competences for EC managers, we proposed to adopt group-decision skills for the confirmation on those items and indicators. Among the group-decision skills, Delphi Method is the one that has usually been applied. The approach is to survey the opinions of experts in the specific fields with questionnaires. Without any

common consensuses or exchanges of ideas among experts, the negotiator will integrate the opinions from the experts and select the average or 50% of the result as the collective opinion. If the result does not satisfy the standard of converge, the survey will need to be repeated several times until the standard is achieved (Hsu, 1990). Therefore, the disadvantages of the conventional Delphi Method are: opinions from the experts might not easily achieve the standard of converge; the survey would often need to be repeated for several times until the acceptable standard is achieved, which might result in high expenses of the capital and time; and with the decrease of response rate, the result might be affected by the negotiator's subjective thoughts (Hsu and Lee, 1997). Hence, if the Fuzzy theory could be applied to the Delphi Method, apart from the benefit that the result obtained could be similar to the conventional Delphi Method, the repeating time for survey could be decreased and the time and capital could therefore be reduced. In particular, the individual characteristic of each expert could be reflected and the professional knowledge of each expert would be more reasonably and suitably dealt with (Ishikawa et al., 1993).

Therefore, our study adopts the reformed Fuzzy Delphi Method introduced by Hsu and Lee (1997), which based on the triangular fuzzy numbers. This method was applied to make confirmation on the managerial competences and aspects for EC managers, because it not only improved the disadvantages resulted from the conventional Delphi Method, but that the result would not easily be affected by the extreme opinions. The questionnaires for the survey on Fuzzy Delphi Methods were sent on 15 June 2002 to the recipients, who were members of EC association in Taiwan, and the experts invited for the focus group. The total numbers of the questionnaires sent were 120 and with 40 valid responses, which made it 33% valid response rate. Finally, the procedure of adopting Fuzzy Delphi Method to the structure of managerial competences for EC managers is described as follows:

Step One: Building the structure of managerial competences for EC managers. Our initial setting of managerial competences for EC managers could be discussed from 8 aspects, which were conceptual skills, social skills, leadership skills, administrative skills, professional skills, personal qualities, abilities and motivations.

Step Two: Setting up the evaluations of weights.

By using the triangular fuzzy numbers to integrate the opinions from experts, the problems of fuzzy human thoughts and inaccuracy could be reduced. Besides, the decision maker's real attempt could be more clarified. The approaches to integrate the group opinions were average number(s), total number(s), the maximum, the minimum, and the hybrid method of

average of the maximum and the minimum. Besides, as pointed by Saaty (1980), it would be better to present the opinions of the experts with the geometric average method, which has been applied frequently in practical studies. Moreover, because the result from the geometric average method would be suitable for the definition of fuzzy judgment matrix, in our study with Fuzzy Delphi Method we used geometric averages to demonstrate the collective opinions from experts. Hence, the triangular fuzzy numbers were adopted, which were the most often applied in fuzzy numbers, for the purpose of setting up the fuzzy membership function of the experts' opinions. In this chart, the minimum of the experts common consensuses is presented as point "l", whereas the maximum of them is showed as point "u". Accordingly, the satisfaction degree to these two extremes would be presented as 0, while the segments between u and l would be given the satisfaction degree between 0 and 1. Hence, the fuzzy number of the total opinions from experts is contained in our study, which is presented as $\tilde{W} = (l, m, u)_{L-R}$, whereas the fuzzy weight number of the indicator for the managerial competence item j by an individual expert i would be $\tilde{W}_{ij} = (l_{ij}, m_{ij}, u_{ij})_{L-R}$. Consequently, the approach described above could prevent from the problem of deleting information as happened in the conventional Delphi Method, and the common consensuses among experts were included as many as possible.

From the survey of Fuzzy Delphi Method, the evaluations of fuzzy weights from the experts were therefore obtained. Hence, the fuzzy weight number of expert i on managerial competence item j could be presented as follows:

$$\tilde{W}_{ij} = (l_{ij}, m_{ij}, u_{ij})_{L-R} \tag{1}$$

$$l_{ij} = \text{Min} (l_{ij}), i = 1, \dots, n; j = 1, \dots, m \tag{2}$$

$$m_{ij} = \left(\prod_{i=1, j=1}^{n, m} m_{ij} \right)^{\frac{1}{n \cdot m}}, i = 1, \dots, n; j = 1, \dots, m \tag{3}$$

$$u_{ij} = \text{Max} (u_{ij}), i = 1, \dots, n; j = 1, \dots, m \tag{4}$$

While n stands for the numbers of experts, m shows the numbers of indicators for managerial competences

Step Three: Selecting the suitable items of managerial competences for EC managers.

1. The fuzzy weight number \tilde{W}_{ij} generated from Step Two could not be used for direct comparison. Therefore, the fuzzy method solution which was provided by Hsu and Lee (1997) was provided in our study, and the result was transformed into a single number of_j .

$$of_j = \frac{u_j + m_j + l_j}{3} \tag{5}$$

2. Then a threshold was provided by experts. From the references of previous studies on managerial

competences, we decided to set the threshold $r=0.7$, and used this for the selection of suitable managerial competences. If $of_i \geq r$, then the managerial competence item should be selected; If $of_i \leq r$, then the managerial competence item should be deleted.

4. DATAANALYSIS

After establishing the initial construct of competences for EC professional managers, a focus group method was used for the purpose of discovering the managerial competences related to the job of EC managers, which also provided some suggestions for the adjustment to the initial structure of our study.

Fuzzy Delphi Method is then applied in our study for the identification of managerial competences and aspects for EC managers. From the three steps for Fuzzy Delphi Method, the initial 73 items of managerial competences in 8 aspects, after deleting 33 less important items, 40 were left. The initial 73 indicators of managerial competences in eight aspects, after the selection with Fuzzy Delphi Method were all kept with 39 managerial competence indicators which have been selected, while the other 34 were considered by experts as unimportant managerial competences indicators and therefore have been deleted. The study result showed that professional skills, conceptual skills, social skills, leadership skills, administration skills and personal qualities were considered by the experts as important aspects. Therefore, most of the managerial competence items in these aspects have been retained; whereas the items deleted were mostly from the aspects of abilities and motivations, which eventually contained only 2 items for each aspect. Hence, the result also suggested that these two aspects were considered by most experts in the survey as unimportant aspects. Accordingly, the result of managerial competences for EC management also showed that viewpoints from the functional school and the investigation school tended to be more supported, while personal qualities and motivations were considered less important in the EC environment.

5. CONCLUSIONS

From the results mentioned above, competencies for EC professional manager can be compared with the study of Chang (1998) on the managerial competences for managers, we could see that managerial competences for EC managers tend to be closer to those for the managers in the intermediate level. For instance, the managerial competences for EC managers correspond to the "professional knowledge" for managers in the basic level; similar to "communication skills", "training and supporting staff members", "incident management and group construction", "innovation" and "negotiation" for managers in the intermediate level; and closed to "strategies planning", "sensitivity to the market", "consultation and authorization" and "prospect" for the managers in the high level. In addition, among the

managerial competences for EC managers, "strategies planning", "problem solving" and "negotiation" also echo the study of Chung (1998) on the managerial competences of higher level managers in the high-tech industry. Hence, we found that in both the business and academic fields in Taiwan, EC managers were considered as the "intermediate" or "high" level managers. One of the reasons for internet bubble economy, as our study suggests, is that the skill aspect had been too much emphasized in the Electronic Commerce. Our study also points out that it would be not enough to concentrate on the skill level only. EC managers still have to possess the thinking, leading and management skills which are in the strategic level. For example, if they could help the enterprise or company build the prospect, plan the strategies and assign the resources; if they are innovative; if they could evaluate and make analysis of the environment and future tendency; if they have good communication skills, incident management skills etc. All these are the managerial competences for a successful EC manager.

From our study, we found that EC management could be divided into eight aspects: 1. professional skills, 2. leadership skills, 3. conceptual skills, 4. abilities, 5. social skills, 6. personal qualities, 7. Administrative skills, and 8. motivations. In these aspects, professional skills and leadership skills are the most important aspects, while Administrative skills and motivations are the least important ones. In terms of the managerial competence items, the main difference between the competences for EC managers and traditional managerial competences lies in the emphasis of professional skills. In addition, domain knowledge and information management have been recognized by the experts as the most important items in professional skills. As for other managerial aspects, such as leadership skills and conceptual skills are similar to traditional management theories, which are also included as managerial competences possessed by successful managers. Besides, in terms of practical usages, the managerial competences and indicators proposed by our study would be very suitable for the companies when evaluating and hiring EC managers. Even for educational trainings, these managerial competences and indicators would certainly provide as a good reference and direction.

ACKNOWLEDGEMENT

This work was supported by the National Science Council of Taiwan, China, under Grand No. NSC-90-2520-S-305-001.

REFERENCES

- [1] Chang, P. T., Huang, L. C., and Lin, H. J., "The fuzzy Delphi method via fuzzy statistics and membership function fitting and an application to the human resources," *Fuzzy Sets and Systems*, Vol.112,

- pp.511-520, 2000.
- [2] Chang, Y. L., "A study of developing an evaluation tool for managerial competence and reliability and validity analysis," *A Research Report of National Science Council* (in Chinese), 1998.
- [3] Chuang, C. C., *Developing Managerial Competencies Assessment Scale-A Case Study in High-Tech Firm*, Master Thesis, Graduated Institute of Human Resource Management, National Central University (in Chinese), 1997.
- [4] Chung, P. T., *A Research on Hi-Tech Industry Managerial Competencies Resources Channel*, Master Thesis, Graduated Institute of Human Resource Management, National Central University (in Chinese), 1998.
- [5] Fielder, F. E., *A Theory of Leadership Effectiveness*. New York:McGraw-Hill, 1967.
- [6] Gattell, R. B., "Personality Pined Down", *Psychology Today*, July, pp.40-46, 1973.
- [7] Guglielmino, "Developing the Top-Level Executive for the 1980's and Beyond", *Training and Development*, Vol.33, pp.12-14, 1979.
- [8] House, R. J., "A path-goal theory of effectiveness", *Administrative Science Quarterly*, Vol. 16, pp.321-338, 1971.
- [9] Hsieh, C. C., *A Prototype of Expert System for Evaluating Managerial Competence: An Example of a High-technology Company*, Master Thesis, Graduated Institute of Human Resource Management, National Central University (in Chinese), 1998.
- [10] Hsu, Paul, S. C., *Management*, Don-Hwa Book Inc, 1990.
- [11] Hsu, T. H. and Lee, T. C., "The Fuzzy Product Portfolio Model", *Journal of Management*, Vol.14, No.2, pp. 263-286, 1997.
- [12] Huang, L. C., Chang, P. T., & Lin, H. J., "The Fuzzy Managerial Talent Assessment Model: A Pilot Study", *Pan-Pacafic Managerial Review*, Vol. 1, No.1, pp.71-83, 1997.
- [13] Ishikawa, A., Amagasa, M., Shiga, T., Tomizawa, G., Tatsuta, R. & Mieno, H., "The Max-Min Delphi Method and Fuzzy Delphi Method via Fuzzy Integration", *Fuzzy Sets and Systems*, pp.55, 1993.
- [14] Katz, R. L., "Skills of an Effective Administrator", *Harvard Business Review*, 33, pp.33-42, 1955.
- [15] Larrhoven, P. J. M. and Pedrycz, W., "A fuzzy extension of Saaty's priority theory," *Fuzzy Sets and Systems*, Vol.11, No.3, pp.229-241, 1983.
- [16] Lee, E. S. and Li, R. L., "Comparison of fuzzy numbers based on the probability measure of fuzzy events," *Computational Mathematics and Applications*, Vol.15, pp.887-896, 1988.
- [17] Lin, H. J., *Applying Fuzzy Theory to An Assessment Model of Managerial Talent*, Master Thesis, Graduated School of Management, I-Shou University (in Chinese), 1996.
- [18] Mansfield, R. S., "Building competency models: Approaches for HR professionals", *Human Resource Management*, Vol.35, No.1, pp.7-18, 1996.
- [19] Maslow, A., *Motivation and Personality*, New York:Harper & Row, 1954.
- [20] McClelland, D., "Testing for Competence rather than for Intelligence," *American Psychologist*, Vol.25, pp.1-24, 1973.
- [21] Murakami, *A book for design and application on personal assessment system*, Tai-hwa book corporate, Taipei, 1988.
- [22] Parry, S. B., "Just what is a Competency? And why should you care?", *Training*, pp.58-64, 1998.
- [23] Richard T. T., and James M. B., "Competencies of Exceptional and Nonexceptional Software Engineers," *Journal of Systems Software*, pp.19-38, 1995.
- [24] Ritter, T. and Gemunden, H. G., "The impact of a company's business strategy on its technological competence, network competence and innovation success," *Journal of Business Research*, Vol.57, pp.548-556, 2004.
- [25] Robbins, S.P., *Organization Behavior: Concepts, Controversies and Application*, Englewood Cliffs, N.J. : Prentice Hall, 1985.
- [26] Sanchez, R., "Understanding competence-based management identifying and managing five models of competence," *Journal of Business Research*, Vol.57, pp.518-532, 2004.
- [27] Scott, B. P., "Just what in a competency?", *Training Minneapolis*, Vol.35, No.6, pp.58-64, 1998.
- [28] Spencer, L. M. & Spencer, S. M., *Competence at Work*, New York: John Wiley & Sons, 1993.
- [29] Teng, G. H. and Tzeng, J. Y., "Transportation Investment Project Selection with Fuzzy Multiobjectives", *Transportation Planning and Technology*, Vol.17, pp.91-112, 1993.
- [30] Tseng, P. L., *A study of the relationships between corporate human resource strategies & core competencies*, Master Thesis, Department of Industrial Education, National Taiwan Normal University (in Chinese), 1998.
- [31] Wu, F. S., "Evaluation and Selection for Manager," *Personnel Monthly*, Vol.17, No.3, pp.21-43 (in Chinese), 1993.
- [32] Wu, Y. H., *A Study of Applying Fuzzy Analytic Hierarchy Process on Management Talent Evaluation Model*, Master Thesis, National Central University, 1999.
- [33] Yeung, A. K., "Competencies for HR Professionals: An Interview with Richard E. Boyatzis", *Human Resource Management*, Vol.35, No.1, pp.119-131, 1996.

A Study of Influencing Traditional Travel Agencies' Decision Making to Introducing B2C E-commerce

Chenyi Hsu¹, Yunghsin Wang²

¹ Graduate School of Management, Tatung University, Taipei, 104, Taiwan, China

² Department of Information Management, Tatung University, Taipei, 104, Taiwan, China
richard_hsu0408@msn.com, ywang@ttu.edu.tw

ABSTRACT

Before the emerging of Internet, traditional travel agencies had been an industry that adopted information technology greatly, and common for each travel agency to have Computerized Reservation System to connect with airlines and hotels through dedicated lines. This kind of inter-organizational information system has been adopted for years; while viewing the travel industry as a whole, the changes on value chain and competition and cooperation relation are caused by the commercial behavior of Internet, E-commerce (EC). Therefore, the objectives of this study are: to understand the influence of top managements' characteristics on introducing B2C EC, to understand the influence of travel agencies' attributes on introducing B2C EC, and to understand the decision-making factors travel agencies considered while introducing B2C EC. To the data analysis, we adopt scale reliability analysis, scale validity analysis, descriptive statistics, chi-square test, independent sample t-test, Pearson's correlation analysis, factor analysis and discriminated analysis, etc. According to the analysis results, this study found that "education level", "job position", "consuming experience in EC websites", and "viewpoint on EC prospective" of the top management's characteristics would directly influence the introduction of e-commerce and have positive correlation with the decision significantly; the travel agency's attributes of "type of travel agency", "company's age", "number of employees" and "participating in the promotion and assistance measure" would directly influence whether to introduce e-commerce or not, and the decision is positively correlated with significance of EC introduction. Decision factors are divided into six dimensions, "Product Competence", "Relationship of Investment Capital", "E-business' Level", "Consumers Demand", "E-commerce Environment", and "Management Knowledge Mode". Results show that the dimension "E-business' Level" affects the most and dimension "Relationship of Investment Capital" affects next the decision to introducing EC.

Keywords: E-commerce, decision-making, B2C EC, travel agencies

1. INTRODUCTION

The public has accepted travel industry as the biggest industry in 21st century which highly interrelated with each person. Traditional travel agency is the service industry that in charge of travel information brokred and management among all travel industry, and the major assets it possessed are customers. Before the emerging of Internet, traditional travel agencies in Taiwan had been an industry that adopted information technology greatly, and common for each travel agency to have Computerized Reservation System (CRS) to connect with airlines and hotels through dedicated lines. This kind of inter-organizational system (IOS) has been adopted for years. To view travel industry as a whole, changes on value chain, competition and cooperation relation are caused by the E-commerce (EC) behavior.

The increase in Internet users, growth in on-line order, open of Internet banking business's payment system operation, gradual mature of relevant environment for e-commerce and popular of on-line games, and markets for e-commerce will keep on growth, and Taiwan's on-line traveling market has been the biggest business among B2C EC now. However, it is true that Taiwan's traditional travel agencies are facing the same problems as most small and medium enterprises, namely, capital, technology, industrial standards and governmental

promotion and assistance measures, which result in the stagnancy of e-business degree. Therefore, the motivation of this study is to discuss and realize how the domestic enterprises' top managers of traditional travel agency look upon B2C EC and what are the required features, and to understand the decision-making factors on introducing B2C EC into traditional travel agencies.

The issues of this research is "What are the top managers' characteristics, travel agencies' attributes, influences of relationship between suppliers and governmental promotion and assistance measures on the decision-making of introducing emerging information technology e-commerce by Taiwan traditional travel agencies?" Therefore, the objectives of this study are:

- (1) To understand the influence of top managements' characteristics on introducing B2C EC.
- (2) To understand the influence of travel agencies' attributes on introducing B2C EC.
- (3) To understand the considered decision-making factors by travel agencies while introducing B2C EC.
- (4) To provide suggestions and references for government's relevant unit, higher education research institutes, system providers and relevant software suppliers, travel product suppliers, travel associations and travel agencies, to construct a well digitalized travel operation ecological environment jointly.

We attempt to use questionnaire survey to discuss how the traditional travel agencies shall find out their ways under this trend and develop business opportunities, and catch the important factors considered under traditional travel agencies' decision-making process, and what their required assistance and help are.

According to business natures, Taiwan travel agencies could be divided into three types, namely, Consolidated, Class A and Class B travel agencies. This study focuses on and sampled from the travel agencies that join the Taoyuan County Travel Agent Association. The amount of travel agencies in Taoyuan area are ranked fourth in Taiwan, and most are the localized traditional travel agencies without on-line traveling operation. In addition, since we are concerning the travel agencies' decision-making to introducing e-commerce and only top management could make such decision, it is more reliable for us to take them as our research objects.

2. LITERATURE REVIEW

This section reviews relevant literatures concerning the decision-making factors that influence the introduction of e-commerce into travel agencies. The following discusses ten of such factors:

- **Top Management's Characteristics:** Subjective factor that influences decision-making rationality is the decision makers, that is, manager's personal intuition, experience and proficiency [3]; McNulin and Sprague [12] considered that a top management with operation styles of perspective, adventure and brave in accepting challenges, also have higher acceptance on the introduction of new information technology; in the research of critical factors that influence top managements' support of information technology from sixty domestic enterprises made by Lin [10], it is found that top managements with higher education background are tended to support information technology due to their higher opportunities to get in touch with information technology.
- **Travel Agency's Attributes:** Scale is also one of the major factors that influence organization's application of information technology; numerous researches believed that organization scale and new technology introduction have positive correlation [1,5,9].
- **Policy Regulations:** With more governmental promotion and assistance measures, the enterprise could even more facilitated to adopt new information technology [6,13].
- **Consumer's Behaviors:** Engel et al. [4] defined consumer behavior as the relevant activities made directly by personal to product, service obtain, consumption or dispose, including causes, practical activity and decision-making process that decide various behaviors.
- **Construction Purpose:** We thought that a success enterprise does not pursue fashion or follow the trend. It must have been driven by certain purpose, which means that enterprise must have profits to obtain and select most appropriate project or technology. Moreover, methods for enterprise to gain profits including improve operation efficiency to reduce operation costs, increase business incomes to accelerate investment return, and expand market to increase new business opportunities.
- **Organization:** Travel agencies' technology maturity, information department's scale and top managements' cognition and support are all the factors that influence the decision-making of adopt information technology into organization; and more formalized company will adopt information technology earlier [1,9,11,14]. There are other researches indicated that if the computerized degree is higher among industry, the organization would correspond with higher formalized organization structure [2,15].
- **Supplier:** Rockart and Short [13] considered that the development of information technology enables the enterprise to link consumers and upper suppliers, and cross-organization's process could closely integrated, which highly increased the interflow of information and obscure the organizational boundaries. The service quality of vertical integration between information system supplier and enterprise will affect enterprise's IT adoption, and more marketing supports and understanding customers' requirements will make products easier to be accepted [9,15].
- **Environment:** (1) Pressure from the adoption of same business: As indicated in Lee [9] and Tseng [14], competition between same businesses, computerized degree between same businesses, and influence degree by the same businesses have positive correlation with the effects of the introduction of information technology; (2) Pressure from customer's demand to adopt: Based on Hwang's research [8] on enterprise's adoption of EDI and its benefits, it is discovered that when more customers request to adopt new information technology, the enterprise could even more facilitated to adopt new information technology; (3) Sporadic environment: According to Grover and Goslar's research [5] of the factors that influence enterprise's application of communication technology in USA, it is discovered that environment uncertainty and heterology would facilitate the adoption of new technology.
- **Internet Techniques:** Include network construction, management and maintenance costs, transmission quality, confidential and safety, integration of various computer systems, network applicable software, and network supplier's service; while these factors have positive development, the enterprise could even more facilitated to adopt new information technology [6,7].

- **Technological Recognition:** (1) Characteristics of Internet: Lee [9] in his study of the Internet adoption decision factor referred that it has different opinions on “whether to adopt or not” in terms of Internet risk; however, the negative factors of compatibility do not affect enterprises’ willingness to adopt Internet; (2) Characteristics of enterprise product/service: Product and service characteristics are unique in travel industry that also reveal their importance. Thus, there are researches addressed factors like product’s comparative profit, comparative risk, compatibility and cognition degree [8,15].

3. RESEARCH METHODOLOGY

This study proposes operational definitions of related variables in accordance with the research hypotheses. The conceptual framework of our research considers the influential factors such as the properties of the travel agency, travel products, environment, the Internet as well as the characteristics of the top management.

- (1) **Top Management’s Characteristics:** The six variables are gender, age, education level, job position, consuming experience in e-commerce websites, and viewpoint on e-commerce prospects.
- (2) **Travel Agency’s Attributes:** It includes type of travel agency, company’s age, number of employees, and degree of participating in e-commerce promotion and assistance measures of government’s relevant units or private organizations.
- (3) **Policy Regulations:** This mainly focuses on the travel agency’s participation in governmental promotion and assistance measures, which might influence the travel agency to introduce e-commerce.
- (4) **Decision Dimensions:**
 - **Consumer’s Behaviors:** This dimension highlights the behavior of the consumers to recognize the opinions of the top management of the travel agency toward what consumers think of e-commerce.
 - **Construction Purposes:** This includes the viewpoints of the top management toward the profit of e-commerce, marketing channel, market share rate and publicity of the company.
 - **Organization:** This includes decision factors such as the policy decision procedure, the document format, the e-business level of the enterprise, the e-commerce technicians, the participation and support of the top management and so on.
 - **Suppliers:** With the characteristics of the travel products and the property of the travel agency, the top management lays an emphasis on the behavior of the suppliers. This is because upstream suppliers are able to positively support the marketing activity and networking supply by sending technicians to assist or providing reasonable preferential prices.

- **Environment:** This includes adoptive pressure from the competitors, customer’s requests, the recognition and viewpoints of the top management on the domestic EC environment, and the consideration of the SARS impact on e-commerce.
- **Internet Techniques:** This aims at the EC establishment costs (implementation, management, maintenance, etc.) and the EC on-line transaction security problems.
- **Technological Recognition:** Whether the Internet is able to sustain marketing, reduce the enterprise operating cost, promote company’s competency, and extend the service district, and e-business is a commercial trend in the future.

The evaluative item of each variable in our research questionnaire is based on the theoretic basis of the literature review and related research subjects, further combined with the research objectives and conceptual framework. This mainly consists of two parts:

- (1) **Questionnaire Design**
 - **Pretest Part:** Having done effective analysis of the questionnaire pretest and deleted questions that don’t meet the criteria, a formal questionnaire is gained.
 - **Formal Questionnaire:** The content of the formal questionnaire altogether is divided into three parts—Basic Information, Operating Status, and Decision Making Factors on E-commerce Introduction.
- (2) **Survey**
 - **Questionnaire Duration:** The survey duration of this research is from June 25th to July 31st, 2003.
 - **Sampling Method:** The research adopts purposive sampling method. The population includes 19 branches of consolidated travel agencies, 95 head offices of Class-A travel agencies, 30 branches of Class-A travel agencies and 7 head offices of Class-B travel agencies, totally 151 travel agencies.

This research adopts SPSS for Windows as a tool for analysis. The flowchart of analysis is shown in Figure 1.

4. DATA ANALYSIS

We now proceed to present data analysis, including chi-square test, independent sample t-test, Pearson’s correlation analysis, factor analysis and discriminated analysis, etc. and explains the results in detail. For this study, 151 questionnaires were sent out and 88 were returned. The response rate was 55.63%. Among those received, 4 were incomplete and treated as invalid.

This study used the Cronbach’s α value to analyze the reliability of scale in part of the questionnaire that concerns the decision-making factors of introducing EC in traditional travel agency. After measuring, the Cronbach’s α value is above 0.94 which satisfies the internal consistency requirement and is highly reliable.

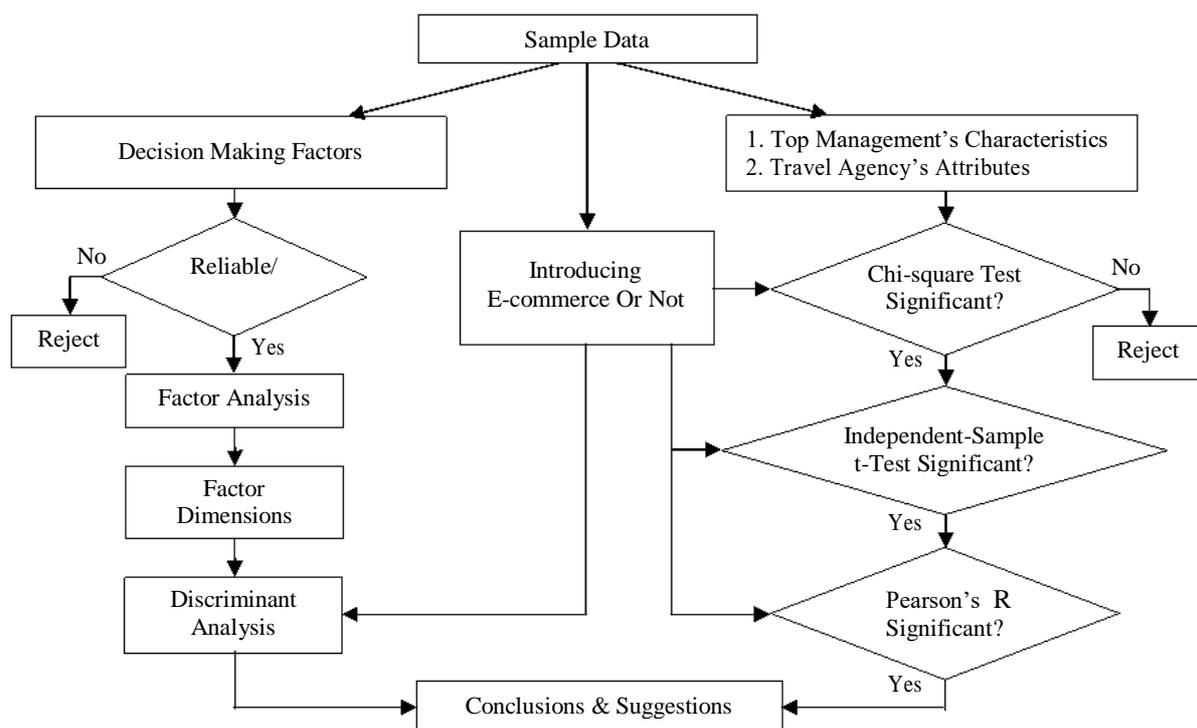


Figure 1. The flowchart of analysis

As shown in Table 1 and Table 2, the education level (Pearson R= .259, p<.01), job position (Pearson R= .254, p<.01), consuming experience (Pearson R= .243, p<.05), and the viewpoint on e-commerce prospective (Pearson R= .222, p<.05) of the top management in the travel agency would have significantly positive influence on the decision making of EC introduction; but the gender ($\chi^2(1)=3.385$, p>.05) and age ($\chi^2(1)=1.911$, p>.05) did not reach significance of confidence interval, so we can say that the gender and age of the top management will not influence the decision making of introducing EC.

Table 3 shows the correlations between travel agency's attributes and introducing EC or not are significant at 99% of confidence interval; so the type of travel agency (Pearson R= .337, p< .01), company's history (Pearson R= .338, p< .01), the number of employees (Pearson R= .369, p< .01) and participating in the promotion and assistance measure (Pearson R= .391, p< .01) of the travel agency's attributes would influence the decision making of introducing EC significantly and positively.

Before analyzing, we first check the correlation matrix between each variance to make sure that the factor analysis is representative. Results show that most related coefficient between variance of each dimension appeared as significance. Then after the KMO and Bartlett's test (KMO statistical value = 0.762 > 0.6; p=0.00 < 0.05), it shows the significance and hence is suitable to do the factor analysis. Clearly, the 26 questions against factor loading are all over 0.5, which is at the acceptable region. For the extracted 6 factor dimensions which has characteristic value bigger than 1, total variance explained is about 75.1%.

Table 1. Chi-Square tests for top management's characteristics on whether to introduce EC

Variables		84 Valid Cases		
		Value	Df	p-value
Top Management's Characteristics (6 Factors)				
Education level	Pearson Chi-Square	5.631	1	.018
	Likelihood Ratio	5.581	1	.018
	Linear-by-Linear Association	5.564	1	.018
Job position	Pearson Chi-Square	5.427	1	.020
	Likelihood Ratio	5.571	1	.018
	Linear-by-Linear Association	5.362	1	.021
Consuming experience in EC websites	Pearson Chi-Square	4.941	1	.026
	Likelihood Ratio	4.999	1	.025
	Linear-by-Linear Association	4.882	1	.027
Viewpoint on e-commerce prospective	Pearson Chi-Square	4.153	1	.042
	Likelihood Ratio	4.095	1	.043
	Linear-by-Linear Association	4.104	1	.043
Gender	Pearson Chi-Square	3.385	1	.066
	Likelihood Ratio	3.344	1	.067
	Linear-by-Linear Association	3.344	1	.067
Age	Pearson Chi-Square	1.911	4	.752
	Likelihood Ratio	2.248	4	.690
	Linear-by-Linear Association	.140	1	.708

Source: This study

Table 2. Correlations between top management’s characteristics and whether to introduce EC

Variables (4 Factors) Top Management’s Characteristics		Introducing e-commerce or not
Education level	Pearson Correlation	.259
	Sig. (2-tailed)	.017*
	Sig. (1-tailed)	.009**
Job position	Pearson Correlation	.254
	Sig. (2-tailed)	.020*
	Sig. (1-tailed)	.010**
Consuming experience in EC websites	Pearson Correlation	.243
	Sig. (2-tailed)	.026*
	Sig. (1-tailed)	.013*
Viewpoint on e-commerce prospective	Pearson Correlation	.222
	Sig. (2-tailed)	.042*
	Sig. (1-tailed)	.021*

Note: *p<.05; **p<.01

Source: This study

Table 3. Correlations between travel agency’s attributes and whether to introduce EC

Variables Travel Agency’s Attributes (4 Factors)		Introducing e-commerce or not
Type of travel agency	Pearson Correlation	.337
	Sig. (2-tailed)	.002*
	Sig. (1-tailed)	.001*
Company’s Age	Pearson Correlation	.338
	Sig. (2-tailed)	.002*
	Sig. (1-tailed)	.001*
The number of employees	Pearson Correlation	.369
	Sig. (2-tailed)	.001*
	Sig. (1-tailed)	.000*
Participating in the promotion and assistance measure	Pearson Correlation	.391
	Sig. (2-tailed)	.000*
	Sig. (1-tailed)	.000*

Note: *Correlation is significant at the 0.01 level

Source: This study

The name of factors was based on variance on the factor loading of factor pattern, and extended to some other related variances. The higher variances, the more important the top management attach to themselves. Thus decision factors are named and classified to 6 dimensions as “Product Competence”, “Relationship of Investment Capital”, “E-business’Level”, “Consumers Demand”, “EC Environment”, and “Management Knowledge Mode”. Then, we implement Cronbach’s a value to prove and measure the reliability of each dimension; if the Cronbach’s a-coefficient is greater than 0.7, it means that the reliability is quite high and the questionnaire is well designed. After factor analysis, the reliability is as shown in Table 4.

This study uses Discriminant Analysis to analyze the effective decision dimension and whether travel agency is to introduce EC. We define “introducing e-commerce factor dimensions” as independent variables and “whether travel agency introducing e-commerce or not” as dependent variable. We shall discuss the decision on

each dimension affecting whether the travel agency introduces EC is significant in reality. Therefore, from the function test of discriminant analysis (Wilks’ Lambda value = 0.772 and p= .000<0.05) which has reached statistical significance level, we can have a standardized canonical discriminant function: D=1.002* E-business’Level + 0.817*Relationship of Investment Capital –0.640*Management Knowledge Mode –0.538 *E-commerce Environment –0.395*Consumer Demand – 0.078*Product Competence. Results (see Table 5) show that the dimension “E-business’Level” affects the most and dimension “Relationship of Investment Capital” affects next the decision to introducing EC. The positive and negative symbol shows the direction of relationship.

Table 4. Reliability scale of the factor dimensions

Factor Dimension	Question No.	a value
Product Competence	5, 6, 7, 8, 19, 20, 21, 23, 25	0.9071
Relationship of Investment Capital	9, 10, 17, 18	0.8746
E-business’Level	13, 14, 15, 16	0.8670
Consumer Demand	1, 2, 3, 4	0.8118
E-commerce Environment	22, 24	0.7431
Management Knowledge Mode	11, 12, 26	0.7007

Source: This study

Table 5. Standardized Canonical discriminant function coefficients

Factor Dimension	Function value (Wilks’ Lambda= .722, p= .000)
Product Competence	-0.78
Relationship of Investment Capital	0.817
E-business’Level	1.002
Consumer Demand	-0.395
E-commerce Environment	-0.538
Management Knowledge Mode	-0.640

Source: This study

5. CONCLUSIONS

This paper finally summarizes and concludes as below:

(1)Structure of Respondents

According to the analysis, we understand that the decision-makers think EC would be a trend in the future that 76.2% of the sample think EC’s future is good and keep on investing it; while 73.8% actively participate in government’s or private organization’s B2B2C EC promotion and assistance measure. In the characteristics of travel agencies, it appears that the scale and history of travel agencies in Taoyuan County is smaller and shorter. The head office of

Class A travel agency is the majority, mainly by regional business, which locally services customers.

(2) Operating Status Analysis of Travel Agencies

- The Internet connection rate of travel agencies in Taoyuan County is beyond 80%, and the best is consolidated travel agencies. In the private website constitution rate, excluding members of "Travel Industry's Data Base System Program" is only 35.1% and it means the pressure of adopting EC is lower in this area. The rate of Class A (27.2%) and B (14.3%) travel agencies' self-own website is low.
- 84 respondents applied EC (including the travel agencies which plan to introduce EC) that takes 59.5%; among them 31% apply EC by self. In the investigation of travel agency's application of EC, most (62%) lack of IT professionals; while 28% of the travel agencies, which don't apply EC, take strategy on internal staff training. On the other hand, 74% of the managers are from internal who generally accept new technology and are familiar with computer. In the application of IT, beyond 90% of respondents use CRS for trading.

(3) Relationships of Whether to Introduce E-Commerce

- In the education level, job position, consuming experience in EC websites, and the viewpoint on EC prospective of the personal characteristics would directly influence whether to introduce EC or not, and their decisions have positive correlation and are significant. But the gender and age of the top management have no related significant correlation.
- The travel agency's attributes of "type of travel agency", "the company's age", "the number of employees" and "participating in the promotion and assistance measure" would directly influence whether to introduce EC or not, and the decision is positively correlated and significant. Travel agencies which participate in promotion and assistance measures of government's relevant unit or private organization's B2B2C e-commerce are much more active in applying e-commerce.

(4) Decision Factors Analysis of Affecting E-commerce Introduction

- To the opinions of the top management in the travel agency, the top 5 important effects on introducing EC ordered by average are: E-commerce is an inevitable business trend in the future, increase company's fame through customer's visit on the EC website, the company is able to understand customers' preferred habitual behavior, increase the interaction between the company and customer, and the additional EC's channel will increase company's market share.
- After the factor analysis, 26 applications for introducing e-commerce decision factors are divided into six dimensions; they are: "Product Competence", "Relationship of Investment Capital", "E-business Level", "Consumers Demand", "E-commerce Environment", and "Management Knowledge Mode". And after the cross validation, the total

predicted rate is beyond 71%.

REFERENCES

- [1] Chen, C.-H., "A Study on the Adoption Process of World Wide Web Home Page By Business Enterprise and Its Effectiveness", Master's Thesis, National Chiao Tung University, Hsinchu, Taiwan, 1996.
- [2] Chen, L.-H., "The Critical Factors for Information Technology Adoption", Master's Thesis, National Sun Yat-sen University, Kaohsiung, Taiwan, 1993.
- [3] Chiu, K.-F., *Enterprise Behavior Science and Decision Theory*, Taipei: Sanmin Publishing Company, 1982.
- [4] Engel, J.F., R.D., Blackwell, P.W., Miniard, Consumer Behavior, 8th ed., Forth Worth: Dryden, 1995.
- [5] Grover, V., M.D., Goslar, "The Initiation, Adoption, and Implementation of Telecommunications Technologies in U.S.", *Journal of Management Information Systems*, Vol. 10, No. 1, pp141-163, Summer, 1993.
- [6] Huang, C.-S., "A Study of the Enterprises Introducing Electronic Data Interchange", Master's Thesis, Tamkang University, Taipei, Taiwan, 1996.
- [7] Huang, M.-Y., "The Relationship between Business Characteristics and the Status of Internet Applications", Master's Thesis, National Chengchi University, Taipei, Taiwan, 1998.
- [8] Hwang, B.-S., "A Study on Using EDI System to the Enterprises and its Benefit Analysis — Taking Customs Brokers for Example", Master's Thesis, Tamkang University, Taipei, Taiwan, 1994.
- [9] Lee, P.-C., "The Study of the Decision Factor of Internet Adoption in Taiwan Business", Master's thesis, Dept. of Information Management, National Central University, Taoyuan, Taiwan, 1996.
- [10] Lin, T.C., "The Research of Critical Factors that Influence Top Managements' Support of Information Technology", *MIS Review*, Vol. 7, pp23-43, 1997.
- [11] Lin, Z.-L., "Adoption Correlates and Share Effects of Internet in Taiwan Organizations", Master's Thesis, Chinese Culture University, Taipei, Taiwan, 1996.
- [12] McNulin, B.C., R.H., Sprague, *Information systems management's new leadership role, Information System Management in Practice*, 2nd ed., Practice-Hall, Inc., 1990.
- [13] Rockart, J.F., J.E., Short, "IT in the 1990s: Managing organizational interdependence", *Sloan Management Review*, pp7-17, Winter, 1989.
- [14] Tseng, C.-M., "A Study on Factors of Adopting Computer Network and Benefit after Implementing", Master's Thesis, Tamkang University, Taipei, Taiwan, 1996.
- [15] Wang, Y.-W., "A Study on Internet Marketing in the Banking Industry", Master's Thesis, Tamkang University, Taipei, Taiwan, 1997.

A Value Chain Based Analysis of Electronic Commerce's Transformation of China's Pharmaceutical Industry

Xiaochen Wang¹, Mingzhi Li²

¹ School of Economics and Management, Tsinghua University, Beijing 100084, China

² School of Economics and Management, Tsinghua University, Beijing 100084, China
{wangxch, limzh}@em.tsinghua.edu.cn

ABSTRACT

In this paper we explore the impact of electronic commerce on the industry structure and market power allocation of China's pharmaceutical industry. Firstly, we describe the overall characteristics of the pharmaceutical industry in general. Secondly, we develop a value chain based framework of analysis for China's pharmaceutical industry. We divide the pharmaceutical industry value chain into three tiers: the initial market, the middle market and the end market. Based on these works, we analyze the profound impacts of Electronic Commerce on the transformation of China's pharmaceutical industry, with particular emphasis on the evolution of market structure and the re-allocation of market power due to the introduction of E-Commerce procedures.

Keywords: pharmaceutical industry, Internet, E-Commerce, value chain

1. INTRODUCTION

In this paper we explore the impact of electronic commerce on the industry structure and market power allocation of China's pharmaceutical industry. Firstly, we will describe the overall characteristics of the pharmaceutical industry in general. Pharmaceutical manufacturers in the market could be divided into two types: pioneer firms and genetic (imitator) firms, and most of the Chinese pharmaceutical manufacturers belong to the second group. As for the pharmaceutical products, after the introduction of *the Pharmaceutical Regulation Rules*, the taxonomy of medical products in China is becoming similar to that of the US, which consists of prescription drugs and over-the-counter drugs. As a result, the market for medical products could generally be divided into prescription market and over-the-counter market.

Secondly, we develop a value chain based framework of analysis for China's pharmaceutical industry. We divide the pharmaceutical industry value chain into three tiers: the initial market, the middle market and the end market. We focus on the situation of the industry structures and market conditions at present. Then in the last part of the paper, we analyze the profound impacts of Electronic Commerce on the transformation of China's pharmaceutical industry, with particular emphasis on the evolution of market structure and the re-allocation of market power due to the introduction of E-Commerce procedures. We conclude that pharmaceutical manufacturing industry will achieve a new stable structure with smaller firm quantity and stronger firm power. In wholesaling industry only a few big firms would exist in the national market. The three-class agent system may disappear. The relative market power of retailers will strengthen in the initial market that may result in monopsony. And The B2C in pharmaceutical retailing industry would pose a chance to re-allocate the

market share between hospitals and pharmacies.

2. THE OVERVIEW OF PHARMACEUTICAL INDUSTRY

Most of the pharmaceutical markets in the world are highly monopolized. In 2000 the top ten pharmaceutical manufacturers' market share was 50% in the world. In comparison with the world market, Chinese pharmaceutical market is highly competitive with a high number of manufacturers and distributors. That's the peculiarity of Chinese pharmaceutical industry, which we will explicate later. Firstly we're looking for some characteristics of the pharmaceutical industry in general.

The pharmaceutical manufacturers in the market could be divided into two types: pioneer and genetic or imitator firms. The pioneer firms are companies who undertake research and development to discover new drugs and bring them to market. They have the patent rights of these new drugs, which ensure them to be the monopoly before the patent expiration. But the R&D and marketing take long time and millions of dollars. That leads to a much higher price of the pioneer firm's product compared with that of the imitators'. Some famous pioneer firms in the world include Merck and Dupont, ect. Only a few of the Chinese pharmaceutical manufacturers belong to this group, such as Tongrentang. The genetic or imitator firms are smaller and don't do much R&D themselves. They produce some "bioequivalent" products to the original branded products after the patent expiration, and their prices are usually much lower. Most of the Chinese pharmaceutical manufacturers belong to this group.

As for the pharmaceutical products, after the introduction of *the Pharmaceutical Regulation Rules*,

the taxonomy of medical products in China is becoming similar to that of the US, which consists of prescription drugs and over-the-counter drugs. The over-the-counter (OTC) drugs could be bought in any pharmacy. Such as vitamins and parts of the antibiotics. This kind of drug's demand is great but not stable, and its profits usually are not very high. The total revenue in the US drug market in 2002 was 214.7 billion dollars, the OTC only took 57.8 billion dollars in it. And the sale of prescription drugs is controlled strictly. The buyers have to hold a prescription by doctors and can only buy the catalog and dose of the drug listed on the prescription. This kind of drug has much more danger for normal people, but its effect is better to the special patients. In fact the prescription drugs take the main market share and have stable demand. As a result, the market for medical products could generally be divided into prescription market and over-the-counter market. One of the most important characters of prescription drug is that the end consumers have less freedom to choose the products at

their will. Instead, doctors or apothecaries make their choices including the category of drugs and also brands within the category. The unique type of principal agent relationship causes problems in the market such as price insensitivity in the retailing market, asymmetric information, agency problem, necessary for regulation etc.

3. A MARKET TAXONOMY OF THE CHINESE PHARMACEUTICAL INDUSTRY AND MARKETS ANALYSIS

3.1 Chinese Pharmaceutical Industry Value Chain

Though the revolutions in Chinese pharmaceutical industry push the industry closer and closer to the international standards. The value chain in China has some special characters. See Figure 1.

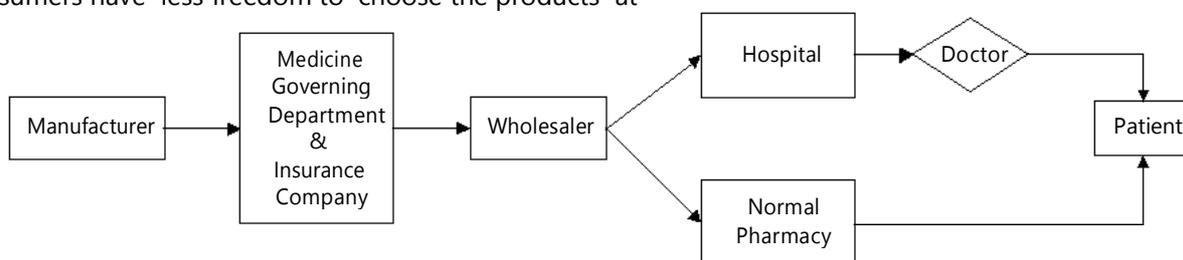


Figure 1 Chinese pharmaceutical industry Value chain

For example, there is no company like the PMB in US. In China the medicine-governing department and insurance companies decide what kind of drugs should be included in the insurance list. Another important feature is that the hospitals also sell medicines as the normal pharmacies in China. So Chinese pharmaceutical retailers include both the pharmacies and hospitals. The hospitals' market share is over 80% in the drug wholesale market. They are the most valuable customers to the wholesalers and manufacturers. In fact, in the prescription market the hospital nearly take almost the whole market share. But in the OTC market, the pharmacy's market share is increasing these years.

3.2 A Market Taxonomy Based on Value Chain — Three-Class Markets

For convenience we divide the value chain into three-class markets and analyze them independently. The buyer in the former class market would turn to be the seller in the latter market. In this paper all the medicine products are thought to be the same (but the prescription and over-the-counter still have some difference which will be mentioned later). We define three-class markets as the initial market, the middle market and the end market, in which the sellers and buyers are manufacturers and wholesalers, wholesalers and retailers, retailers and consumers.

3.3 The Analysis of the Three Markets

3.3.1 The Initial Market

This market is the first market of pharmaceutical products. The manufacturers are the sellers, and the buyers are wholesalers.

At present there are more than 6000 pharmaceutical manufacturers in China. But most of them produce only one or several main products. The average firm size is very small. The biggest one's sale revenue is equal to 1% of the sale revenue in US. The isolation of different regions market is very notable. The firms are competing intensely and there's no real industry leader till now. Another feature is the manufacturers simply focus on production but don't pay much attention to the distribution channel. So they highly rely on the downstream firms who handle the market channel. That's why their market power is weak facing the distributors and agents.

We subdivide the manufacturers into four types and observe the top ten firms' market share from 1996-2002 (See Figure 2). It's clear that the market concentration rate is low. In 2002, the four subdivided markets' CR are 44%, 32%, 18% and 16%. Considering the fact that few of the top ten firms in the four subdivided types are inter-industry. The market concentration of the whole pharmaceutical manufacture is much lower.

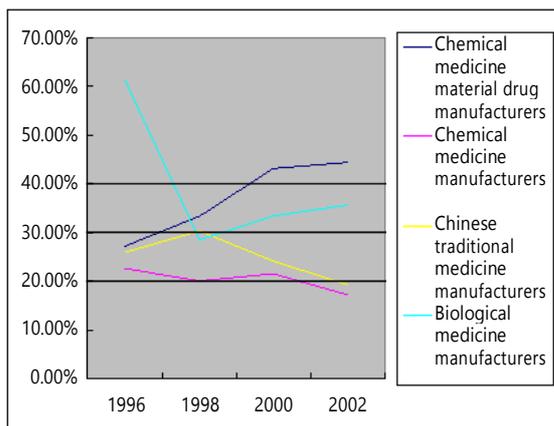


Figure 2 The total market shares of the top ten pharmaceutical manufacturers

As for the wholesalers, the situation is similar. Now there are more than 16000 pharmaceutical wholesalers in China. Only 5% of them have got the revenue per year over ¥ 20,000,000. In 2002, there are 17 pharmaceutical wholesalers on the list of China top 500 wholesalers. The total market share of the 17 firm is 25.88%, CR₁₀=18.98%. Compared with the situation in the US and Germany, the market concentration is very low. See Figure3

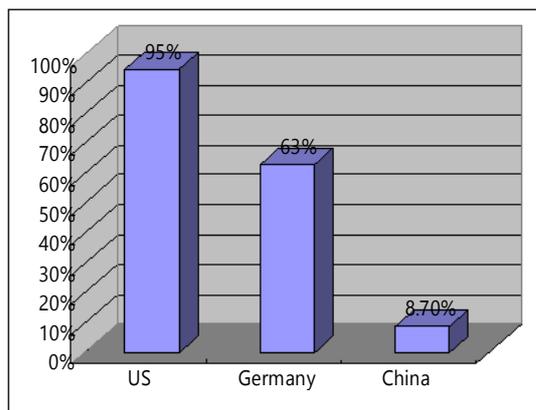


Figure 3 The CR₅s of three countries' pharmaceutical wholesale industries

To sum up, in the initial market there are a great number of manufacturers and most of them have little market power which leads to a low market concentration. The barriers to entry are actually lower than in some other mature product markets. Most of Chinese firms are generic firms whose R&D inputs and levels are relatively low. To the wholesalers who are the buyers in the initial market, there are low barriers to entry. In such a situation both the buyers and sellers do not have great bargain power. But in the regional markets some wholesalers have relatively high monopoly power due to the isolation of the markets from competition.

3.3.2 The Middle Market

The wholesalers are sellers in this market. The retailers

who are the buyers including two groups: pharmacies and hospitals. In theory the two kinds of retailers are the same, but actually the hospitals have obvious advantage, especially in the prescription drug market. They are the most important customers to the wholesalers. This tier of market has been very inefficient because of the wholesalers' iniquitous performance.

But the situation is changing in these years due to the newly adopted policy of the hospital procurement bidding system. Though this bidding system has not been fledged and adopted by all the firms in the industry. It has shown a very attractive tendency. It plans to take all the manufacturers, wholesalers and retailers together in one system including E-Commerce system and EDI. All the information about the drugs such as price, producer, agent or wholesalers will be supported in the E-Commerce system. The retailers would be able to choose the catalog, brand, and the wholesaler as the supplier of one product directly and efficiently. And the wholesalers and retailers could search and exchange the information through EDI. Now the government is trying to develop and propel this pharmaceutical industry electronic business system.

3.3.3 The end market

The sellers and buyers in this market are retailers and consumers. The total demand of the pharmaceutical products is very stable and has low price elasticity. The agent problem is very marketable in this market. That gives the hospital an unshakable advantage.

The demand of every consumer depends on the condition that the consumer is in. If the consumer is in healthy condition, his demand elasticity is zero but the demand is zero also. If the consumer is in ill condition, the demand isn't zero with a very small elasticity. We infer it's a concave function in a limited area whose slop changes from zero to infinity. So the end consumer's demand function could be written as a random variable $D(\omega)$, $\omega \in (\omega_0, \omega_1)$, $\omega_0 = (\text{Healthy})$, $\omega_1 = (\text{Sick})$, that's the most simple space and the could be subdivided farther. Set P as price, Q as quantity, we have

$$D(\omega_0) = 0 \tag{1}$$

$$D(\omega_1) = Q(P) \quad \text{with } Q'(P) < 0 \text{ and } |Q''(P)| < 0. \tag{2}$$

4 APPLICATION OF E-COMMERCE IN THE CHINESE PHARMACEUTICAL INDUSTRY

Now the E-Commerce of pharmaceutical industry is just at the beginning in China. The trading amount is a quite small proportion of the big market. E-Commerce emerges mainly in B2B model. And there hasn't been an efficient regulation to this new model of business. But it's growing up in a remarkable speed. More and more companies have recognized the coming revolution from the new market factor and are going to join the process whether they wish it or not. With the government's effort, there is no doubt that the Internet will be a crucial

factor of the revolution of pharmaceutical industry in China.

In China some successful B2B models are coming in sight. Some of them try to build an information center and industry database such as www.cpi.ac.cn, www.gm.net.cn. Some of them focus on the E-commerce Service Provide (ESP) like www.e135.com, yy.18-china.com. The main customs of these businesses are pharmaceutical manufacturers, wholesalers or agents, pharmacies and hospitals. The most important investment comes from the relevant departments in government and companies. Though their market share is not high, the potentiality of them is very large. Such as www.e135.com, the trading amount in the first 8 months went beyond 15 million Yuan. The amount at the end of 2000 had reached hundreds of million Yuan.

5 THE PROFOUND IMPACT ON CHINESE PHARMACEUTICAL INDUSTRY OF E-COMMERCE: THE INDUSTRY AND MARKET POWER

The E-Commerce of pharmaceutical industry, especially the B2B, is being adopted and going to be the new focus of the Chinese E-Commerce with the support of government. B2C, which has been popular in US, is also a hot point in China. These new business models are the most immediate impacts of the Internet.

Relatively to the B2C, B2B is more practical in the initial market and middle market and may break down the original market structure. The Chinese market is has some remote regions that are inaccessible to the Internet. The barriers to entry of the regional markets are becoming lower, which will lead to emergence of more enters. So the regional monopoly will face more potential competitors and the small firms will have the ability to enter more markets. The result is the changes of the industry structure and the market power allocation of both upstream and downstream firms.

5.1 Manufacture Industry

In the manufacture industry the beneficiary would go to the middle firms. The large firms have paid great cost to get the national market share during the past time. But with the Internet the other smaller firms could get into the national market in a much cheaper way. That's impossible in the no-Internet days. So their prices could be much lower than that of the large firms', which makes them to be serious competitors to the large firms. Though the large firms could apply the Internet to decrease their cost too. The past cost and huge traditional market nets will be their disadvantages in the Internet time. There will not only be the finance problem but also the management problems transformation. And there don't exit such problems for the middle and small firms. They only get a new

competitive advantage from E-Commerce. That may lead to a revolution in the industry.

But what's the problem with the small firms? Why we say the winners should be only the middle ones? That's the special situation in China. Now all the pharmaceutical manufacturers are facing the GMP acceptance. To reach this level the firms have to pay much cost to improve the production conditions, otherwise they will have to quit. And in the new hospital procurement bidding system, some firms are usually not the chosen ones. Based on all the policies of recent years, we infer that the government wants to push the small firms who do not have competitive advantages out of this too-crowded industry. That's the first step to solve the problems in the ex-competitive pharmaceutical industry. The new competitive advantage supported by Internet can't fight with the government. Additionally the small firms in China are highly relied on the local markets. But the Internet will break the local market and bring all the middle and large firms into every regional market. That will be another hit to the small firms. So the winners would only be the middle firms who could catch this chance.

To conclude, The application of B2B in pharmaceutical industry will lead to a new structure: the small firms without competitive advantage will quit the market; some middle firms will have great development by this chance; the big firms have to undergo some revolution to deal with the new competitors to protect their market shares. More combinations will emerge for expansion, some of them will go beyond the regions.

5.2 Wholesaling Industry

In the retailing industry the Internet breaks down the limit of regions that is two-way relatively. That means in both the upstream and downstream markets the wholesalers faces the region limitation. This two-way region limitation may be one important reason for the low market concentration rate.

Comparing with the manufacture industry, the focus in retailing industry is the distribution channels. The Internet will break all the market bounders. That means every wholesaler will face incredible numbers of downstream retailers, which is impossible in the traditional market. Then every firm will have chances to win some new customers and lose some old costumers. That will lead to a higher competition. But the government will play an important role here. Similarly to the manufacture industry, new hospital procurement bidding system will help the powerful firms get more market shares in the future, even result in nature monopoly. There will be some problems of this method, such as the region limitation in retailing market. But the Internet would help overcome it. All the firms who win the bidding will come to one national market by the Internet and compete with each other. So from both

the policy and the technology perspectives, we suggest and prescription drugs respectively. that the government contracts the market and the Internet connects the market.

Acknowledgement

5.3 Retailing Industry

The situation in retailing industry is a little different because the region is not the main factor to limit their development. And they also use the B2C model that has obvious advantage compared with the brick and mortar shops. The e-shop is a real sense 24 hours pharmacy and lower the consumers' supply cost and traffic cost. For the retailers the e-shop has no fixed cost and little running cost. The most attractive point is that it contains a no-inventory way during keeping a proper relationship with suppliers. Therefore in the future this B2C model must become a new focus in retailing market and expands the pharmacy's market share.

6 CONCLUSIONS: SOME PREDICT TO THE FUTURE OF CHINESE PHARMACEUTICAL INDUSTRY

At last the pharmaceutical manufacturing industry will achieve a new stable structure with smaller number of firms and stronger firm power that will lead to a higher market concentration. The market won't be divided into many regional markets with more beyond-regions companies who will compete together in the national market. Comparing with before, the manufacturers' market power will strengthen, but the downstream firms change also, the relative market power depends on both sides. In the wholesaling industry only several big firms would exist in the national market. The three-class agent system may disappear. The relative market power of retailers will strengthen a lot in the initial market that may result in a monopsony situation. The B2C business model in the pharmaceutical retailing industry denotes a chance to re-allocate the market share between hospitals and pharmacies. The final structures may be the pharmacies and hospitals handle the over-the-counter

The research is funded by the National Natural Science Foundation of China (Project Number 70231010 and 70321001).

REFERENCES

- [1] David Waterman, "Local monopsony and free riders", *Information Economics and Policy*, Volume: 8, Issue: 4, pp337-355, 1996.
- [2] Fiona M. Scott Morton, "Barriers to entry, brand advertising, and generic entry in the US pharmaceutical industry", *International Journal of Industrial Organization*, Volume: 18, Issue: 7, pp 1085-1104, 2000.
- [3] Professor Richard L. Nolan, "Drugstore.com", *Harvard Business School*, 9-300-036, 2000.
- [4] Michael E. Porter, "Strategy and the Internet", HBR OnPoint Article, Product Number: 6358, 2001.
- [5] Lei Shi, Zonglai, Kou, *Industry Economics*, Vol. 1, Shanghai Sanlian Publishing, 2003.
- [6] Gu Mi, Wang Yong, Ning Xianfeng, Xu Dongdi, Chen Jieyu, "Analysis of Chinese Pharmaceutical Market", School of Economics and Management in Tsinghua University, 2004.
- [7] *Chinese Economics Yearbook*, 1997-1999 and 2003.
- [8] *Chinese Market Yearbook*, 1999-2004.
- [9] *Chinese Market Statistic Yearbook*, 2003.
- [10] *Chinese Statistic Yearbook*, 1994-2003.

An Exploratory Study of the Electronic Distribution of Hotel Rooms in Europe

David C Gilbert, Liz Lee-Kelley¹, Dominic W Beveridge²

¹School of Management, University of Surrey, Guildford, Surrey, GU2 7XH.

²Cap Gemini Ernst & Young, Sydney, Australia

ABSTRACT

Hotels sell a great deal of their inventory by means of third party distribution through the use of technology and marketing intermediaries. More recently the Internet along with other emerging technologies has offered the potential to disrupt enduring distribution arrangements. The question of strategic change is important as the low-cost, multi-channel possibilities offer considerable potential benefits to hotels, but these are also linked to considerable risks. The focus of the study is on the drivers behind the choice of distribution intermediaries by hotel groups. A research questionnaire designed to establish the factors influencing hotels' choice of intermediaries, was developed from the extant literature. Subsequently principal component analysis uncovered an overriding factor, (referred to here as 'risk preference') which is shown to influence heavily channel strategy choice. This indicates the concern by hoteliers over the impact of negative direct or intermediary-related performance on hotel brand image and reputation. The results of this study further support the theory that the hotel industry has yet to define a strategic direction that will leverage the capabilities of the Internet.

Keywords: hotel distribution, Internet, channel choice

1. LITERATURE REVIEW

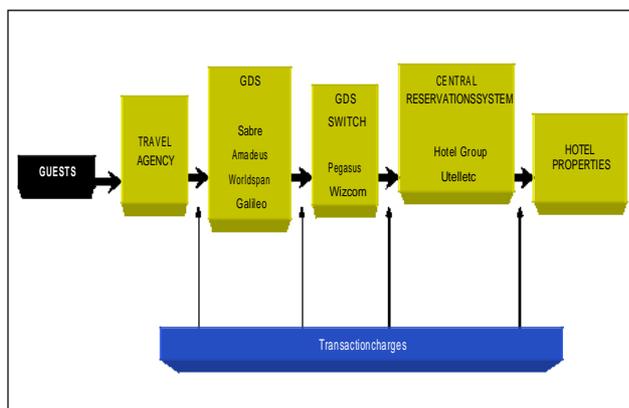


Figure1: The traditional path of online hotel reservations

Source: Price Waterhouse Coopers (1999). *Hospitality Directions European Edition - The Internet Transforms the Traditional Hotel Distribution System*, London, England: Price Waterhouse Coopers.

Distribution systems extend the number opportunities for a sale and facilitate the purchase of products in advance of their production. A traditional distribution path is illustrated in Figure 1 above.

The role of the various European travel agents is to act as the intermediary that brings travel products to the consumer. According to Dube and Renaghan [2] travel agents and meeting planners book almost half of all hotel reservations. As travel industry consumers normally cannot see the product at the time of reservation - agents therefore fulfil the key roles of advising customers, passing information between the different parties, and providing value-added services. Hence, travel agencies by tradition have played an important distributor role - their main advantage is that they are able to reach a larger number of potential

customers than if the provider were to distribute through their own outlets. As demonstrated in Figure 1, numerous intermediaries perform various functions along the transaction chain between the hotel guests and the hoteliers which result in transaction charges.

Global distribution systems (GDS) such as Sabre and Amadeus have been effective means of communication and inventory exchange between local, national and global suppliers, and the travel agencies. GDS accounts for about a third of all hotel bookings world-wide, and bookings are continuing to grow. Corporate travel agents have heavily influenced the development of the GDSs given they are able to control the relationship between the corporations they serve and the suppliers. The corporations required a global network of products which in turn required ownership of global databases, which meant the suppliers had to depend heavily on GDSs for distribution [9]. Another form of intermediary is the switch companies placed between the GDSs and the hotel central reservation systems (CRS). The need for switch companies exists because the wide variety of software and programming languages used by hotel reservation systems must be adapted to GDS software and vice versa [10]. Interestingly, these Switch companies are now able to communicate with consumers over consumer-oriented websites [7]. The CRS have enabled hotels to maintain a single image of inventory between their own systems, and (via the switch), the GDSs. Although the GDS channels attract an increasing volume of hotel reservations, it should be remembered that the bulk of world-wide hotel reservations are made directly with properties or with Central Reservations Office [6]. Although these channels account for a significant volume, they have high administrative costs attached. They both require toll-free telephone numbers in major markets, and trained multilingual staff to handle enquiries. There is evidence to suggest that hotels have been slower than

other industries to realign their distribution strategy towards Internet channels. For example, according to Harteveidt et al [3], in 2000 more than 75% of American online travel bookers bought air tickets, but only 59% booked hotels.

A new form of cyber-intermediary is the online consolidator who negotiates discounts with suppliers and then sells that capacity to consumers. Typically interacting with individual hotels, consolidators act as room brokers that purchase inventory at a deeply discounted rate, selling it on at a mark-up [11]. These online agencies (e.g. Travelocity, Microsoft Expedia and Internet Travel Network) pose a real challenge to the hotel providers since they are unable to match the intermediaries' experience and product/service range, leading O'Conner [6] to conclude that future growth will be driven through online agencies.

Webb [13] advises that hotels should employ a variety of online distribution channels as the benefits of a multi-channel strategy are highly applicable to hotel products:

- Greater responsiveness to customer shopping patterns
- Decreased dependence on existing channels (a single channel may not be suitable for all products)
- Increased outlets for distressed inventory
- Ability to penetrate new markets

From the consumer's perspective, despite the obvious attraction of the convenience of purchasing online, there has been a persistent concern about security, privacy, service levels and trustworthiness. It is also difficult to quantify value added by internet channels when many consumers still employ them purely as a means of pricing rooms perhaps because of the difficulty of physically checking products/services before purchase and the helplessness felt in offering credit card details during online purchase.

O'Connor [6] carried out a Delphi study amongst a group of experts in the field of electronic distribution to develop, validate and weight a list of core evaluation factors for electronic intermediaries. The research was focused on identifying evaluation processes, and it found that there is currently no widespread agreement in the industry on how channels should be evaluated prior to adoption. Financial factors were the most frequently cited group, while technical and operational factors received the highest individual scores - suggesting a need for prioritisation of factors in the evaluation process. Performance appraisal activities tend to be intuitive rather than based on formal guidelines or principles. A possible weakness in this study is its focus on corporate electronic distribution specialists, rather than including hotel-level operators.

2. RESEARCH METHODOLOGY

This study attempts to overcome the complaint in a recent paper [8] that indicated little is understood about

how electronic distribution studies might be evaluated. The research was carried out through a survey (see appendix) that was sent to a sample of key hotel operators with responsibility for channel management. The survey was based on a set of questions designed to measure attitudes towards twenty key attributes of online distribution channels. The responses to the survey are subjected to factor analysis to identify the underlying factors that influence their attitudes. The data collection phase was done by presenting a list of questions about the criteria used in evaluating electronic distribution media by hotel groups. The first task, therefore, was to establish a comprehensive and detailed set of decision criteria that would be assembled for this survey. An initial list of criteria was produced based on the views represented in the literature reviewed. This list was distributed to a reference group of knowledgeable academics, and their feedback was incorporated into a revised list. The revised list was then distributed to a group of five hotel industry e-commerce specialists –all of whom hold management positions in major hotel chains. Their feedback was reviewed and collated to produce a final list of items – these are listed in Table 1. This pool of twenty items formed the basis for the questionnaire.

The questionnaire investigated attitudes and used five point Likert Scales. A mixture of positive and negative statements was presented in the questionnaire to improve accuracy of response. For data collection purposes two separate populations were defined relating to the two sets of target respondents –individual hotels, and corporate level management with responsibility for distribution. The study was carried out with assistance from a number of major international hotel groups. In total, a base of 872 individual hotels was contacted for the study, and a survey instrument was distributed to every hotel in that database. The target recipients were the key rooms' management contacts in each property. A population of corporate hotel groups was also defined using assistance from industry contacts, and the survey was distributed to 70 distribution/channel managers from hotel groups in Europe. The questionnaires were distributed by email containing a link to a website, upon which a collection page was created. The responses from the website were forwarded by Email to a dedicated account from which the data were transferred to a database. The questionnaires were also sent as an attachment to the email, for those respondents with access to Email, but not Internet. Responses sent by email were added manually to the same database. From the 872 individual hotel contacts a total of 268 responses were received, representing a response rate of 30.73%. Of the corporate level respondents, 47 responses were received, representing a response rate of 67.1%. The total combined size of the sample was therefore 315.

Table 1: Evaluation criteria for electronic distribution intermediaries

Subject	Item
Technical capabilities of intermediary system	Method of reservations delivery
	Connectivity to the existing reservation systems
	Availability of booking data
	Security
	Hotel's ability to control price
	Ability to support revenue management strategy
Costs	Start-up/one-off charges for participation
	Commission costs
	Other transaction (eg switch, CRS)
	Processing (eg fax/voice reservations)
Supplier profile	Ability to reach specific markets
	Potential production volumes
	Ownership of medium
	Credibility of owners
User-facing capabilities	Speed and reliability
	Trust/privacy

	Customer experience
Representation of hotels	Appropriateness of site for the hotel's product
	Representation of price to customer
	Content (eg hotel-specific information capabilities)

3. RESEARCH RESULTS

The data collection exercise centred on a list of twenty questions that were administered to respondents. For this analysis, the responses to both the hotel and corporate-level questionnaires were combined to form a single data set of 315 cases, and subjected to factor analysis. The object of this phase of the analysis was to identify the dimensions or factors that are tapped by those twenty variables. The form of factor analysis used, Principal Component Analysis, achieves this by reducing the total number of variables, retaining only those factors that explain a significant proportion of the variance in the overall dataset. The result of the analysis has been to reduce the overall data set to six factors that explain the majority of the overall variance.

Table 2: Summary of factors emerging from analysis

Factor	* % Variance	Items	Subject	Loading	Alpha
1	15.00%	Q3	Security of connection to hotel CRS	0.62	0.6884
		Q12	Customers' trust and privacy	0.70	
		Q13	Appropriateness of site for hotel's image	0.68	
		Q19	Credibility of intermediary owners	0.73	
2	9.82%	Q7	Method of reservation transaction delivery	0.68	0.5733
		Q16	Ability to reach specific markets	0.80	
3	9.60%	Q9	Potential production volumes	0.61	0.4768
		Q14	Transaction charges (eg GDS, switch)	0.71	
		Q18	Hotel's ability to control price	0.57	
4	9.14%	Q4	Availability of management information on bookings	0.70	0.5068
		Q11	Ownership of the intermediary	0.79	
5	8.43%	Q5	Representation of prices	0.70	0.5162
		Q15	Speed and reliability of website	0.76	
6	7.87%	Q2	Connectivity to existing reservations systems	0.81	0.4254
		Q10	Commission rates	0.70	

* Rotation Sums of Squared Loadings

The reliability of the factor analysis results was tested using Cronbach's alpha. The six factors extracted by the analysis are detailed in Table 1, along with the items from the questionnaire that loaded most heavily upon them. It should be noted that the analysis appears to

have identified factors quite different to those suggested by the academic and industry expert reference groups (see Table 1). Based on the responses to direct questions from a sample of industry operators, the logical groupings listed in Table 1 appear not to influence

individuals' attitudes in the way that was originally proposed. It is important, therefore, to understand the factors that have been identified by this analysis. The first factor accounts for a significantly larger proportion of the variance in responses than the others, and an alpha value of 0.69 suggests that it is a relatively reliable measure. Interestingly, the four items that load most heavily on this factor are drawn from four of the different groups defined in Table 1, meaning that this factor is not at all consistent with the groupings defined during the preparatory research phase. This does not mean, however, that the combination of these attributes does not make sense.

Security of connection and customer trust and privacy can be defined as hygiene factors –i.e. they are perhaps unlikely to be identified as a key source of competitive advantage, but the risk of failure in either area is generally unacceptable. The appropriateness for the hotel's image and the credibility of intermediary owners also load heavily on this factor, but like the others, the importance of these items is primarily associated with risk – in particular the damage that appearance on an inappropriate site may do to a hotel's brand image, and the risk of engaging in a partnership with a supplier that results in failure. In short, this factor appears to suggest risk has considerable influence on respondents. Of the seven factors, this was the only one that combined the internal reliability of scale with a plausible explanation. Although several other factors emerged, these had relatively lower reliability scores, and contained items that appeared to be unrelated. There are two possible explanations for this: Firstly the research instrument may not have been interpreted as consistently as the pilot phase suggested. The other explanation is that the approach of respondents to electronic distribution media is genuinely not governed by consistent, logical processes. This finding supports the literature review that suggested a lack of understanding of channel strategy in the hospitality industry.

4. CONCLUSION AND RECOMMENDATIONS

Of the twenty questions in the survey, six relate to the capabilities and the costs of different intermediaries. Had any combination of these items loaded heavily on a single factor, a preference for particular attributes of intermediary offerings would have been indicated. This was not the case, so this survey has not provided such weighting of any future direction regarding online hotel reservations.

Much of the benefit of disintermediation identified in the literature review is associated with cost reduction [1, 3, 4]. The results of the survey, however, offer no clear indication of the importance of different costs in the channel management decision making process. This finding is consistent with other sources add that the cost benefits of online intermediaries are not being realised by hotels. The lack of clear factors influencing

respondents may be explained by the diversity of different job roles amongst the respondents. The survey was distributed to "Key Rooms Management Contacts". This may have impacted the research, although the appropriate contact for each hotel may have been correctly identified, the level of involvement of each individual in both the management of distribution channels, and in the overall strategy of hotels may vary significantly.

Solomon et al [12] further define involvement as the motivation to process information, suggesting that there is a perceived linkage between the user's goals, values, needs, and their product knowledge. It can therefore be argued that the nature of different job roles within the respondents to this survey may explain the apparent lack of factors that clearly dominate the responses. A hotel General Manager, for example, may have ultimate responsibility for their hotels' strategy; however, his or her level of familiarity with the practice of channel management may be very limited if the execution of the task is delegated within the hotel. Conversely, if the key contact is a Reservations Manager, he or she may be highly familiar with the operational procedures of channel management, but not ultimately accountable for the profitability of the hotel whose inventory they manage. Similarly, distribution and channel management positions at corporate level in hotels cannot be regarded as uniform throughout all companies as the roles, responsibilities and the seniority associated with a particular job title varies between hotel groups.

Although only one main factor emerged from the analysis, the overall results of the survey suggest that the opportunities offered by the Internet are not fully realised by the hotel industry at present. This is consistent with many of the views represented in the literature review [5, 3, 6]. Because the study focused on a B2B business model, the findings have implications for management of two groups of businesses: The hotel groups whose inventory is sold through the distribution channels and the intermediaries that provide the distribution services. For hotel operators an important observation from the results of the factor analysis (See Table 2) is that they bear no resemblance to the groupings of factors that were defined by industry experts before the research began. The items were grouped into: Technical capabilities of the intermediary system, costs, supplier profile, user-facing capabilities, and the representation of hotels. These groups appear to be based on entirely logical relationships; however, the factor analysis does not identify any of these groups as underlying psychological dimensions influencing the responses to the questionnaire.

For hotel groups, for example, the importance of cost cannot be understated –The literature review explains the relevance of channels of distribution to hotels, citing the efficiencies that can be realised through the use of distributors. However, none of the factors emerging

from the data grouped responses to the questions about distribution costs together. While this could be attributed to limitations in the research technique, it could also form the basis of the argument that the cost implications of distribution channels are not fully understood by those responsible for executing channel strategy. This is unsurprising – the definition and attribution of costs in hotel distribution is a complex matter. In addition to the structure of charges illustrated in Figure 1, the one-off set-up costs of the intermediary must also be considered – already this suggests a structure of fixed and direct costs. The existence of online consolidators who sell deeply discounted hotel inventory at a mark-up further complicates the issue. The margin between each intermediary's selling and buying prices for each hotel room night is an opportunity cost to the hotels, even though it is seldom if ever treated as a cost item by hotels. Although the complexity of distribution costs makes this a difficult area to understand, it should be a critical factor in channel strategy. The results of the survey suggest that currently it may not be.

Having identified that risk has a strong influence, suppliers could adjust their marketing communications and positioning to emphasise such qualities as the security both of customer details, and of connections to hotel reservations systems, and the image of their site and the owners' financial credibility. It could also be concluded that hotel groups appear not to have a consistent perception of the value associated with electronic distribution intermediary offerings. This view is supported by the survey responses, and it would logically support the recommendation that suppliers have not been effective in communicating the value proposition of their offerings to the marketplace. The sources in the literature review suggest that although technologies and intermediaries that disrupt the traditional reservations chain are now widely available, hotels have been slower than other travel-related industries to embrace the opportunities they offer. Perhaps the opportunity remains for intermediaries to educate the market in the importance of a clear channel strategy, and the implications of different channel offerings for different businesses.

REFERENCES

- [1] Connolly, D. J., "Understanding Information Technology Investment Decision-Making". In *The Context of Hotel Global Distribution Systems: A Multiple-Case Study* Available from: <http://scholar.lib.vt.edu/theses/available/etd-113099-200845/>, 1999.
- [2] Dube, L., & Renaghan, L. M. "Marketing Your Hotel to and through Intermediaries", *Cornell Hotel and Restaurant Administration Quarterly*, Vol. 41, No.1, pp73-83, 2000.
- [3] Hartevelde, H., McCarthy, J. C., DeMoulin, G., & Wakeman, M. "Lodging Online: Rooms to Fill", *Forrester Research Brief*, San Francisco, USA: Forrester Research Inc., 2000.
- [4] Hartevelde, H., McQuivey, H., & Denton, A. "Web Travel Agencies Set to Become Marketmakers", San Francisco, USA: *Forrester Research Brief*, Forrester Research Inc., 2002.
- [5] O'Connor, P. *Electronic Information Distribution in Tourism and Hospitality*, Oxford, England: CABI Publishing, 1999.
- [6] O'Connor, P. *Evaluating Hotel Electronic Channels of Distribution: A Corporate Perspective* Edinburgh, Scotland: Queen Margaret University College, 2001.
- [7] O'Connor, P. & Frew, A. "The Future of Hotel Electronic Distribution", *Cornell Hotel and Restaurant Administration Quarterly*, June, pp33-45, 2002.
- [8] O'Connor, P. & Frew, A. "An Evaluation Methodology for Hotel Electronic Channels of Distribution", *International Journal of Hospitality Management*, Vol. 23, No. 2, pp179-199, 2004.
- [9] PhoCus Wright, *The PhoCusWright Report Online Corporate Travel 2001-2002 –Market Share, Strategy, Forecasts and Commentary*, CT, USA: PhoCusWright, 2002.
- [10] Price Waterhouse Coopers, "Hospitality Directions European Edition - The Internet Transforms the Traditional Hotel Distribution System", London, England: Price Waterhouse Coopers, 1999.
- [11] Rich, M. "The Internet Introduced A Middleman To The Hotel Industry - Now Hotels Want That Business Back", *Wall Street Journal*: R10 Monday 10 June, 2002.
- [12] Solomon, M., Bamossy, G., & Askegaard, S., *Consumer Behaviour: A European Perspective*, Essex, England: Pearson Education Limited, 1999.
- [13] Webb, K. L., "Managing Distribution in the Age of Electronic Commerce", *Industrial Marketing Management*, Vol. 31, No. 2, pp95-102, 2002.

[1] Connolly, D. J., "Understanding Information Technology Investment Decision-Making". In *The*

Appendix: Hotel Electronic Distribution Questionnaire

This questionnaire is part of a study into the electronic channels through which hotel rooms are sold. All responses are entirely confidential. Please answer all of these questions openly.

	<i>Please respond to each statement by entering "X" under one number on the scale. The scale runs from 5, that indicates "I strongly agree" through to 1, that indicates "I strongly disagree".</i>	I strongly agree			I strongly disagree	
		5	4	3	2	1
1	I only agree to distribute through websites that provide bookers with good and detailed information pages on hotels.					
2	I only look to sell through channels that can connect directly to my hotel's central reservations system.					
3	I would not consider connecting my hotel to a distribution channel unless I could be sure that it is secure.					
4	It is essential for booking channels to provide management information about bookings.					
5	Before agreeing to sell rooms through an electronic channel, the way that prices are represented must be made clear.					
6	The cost of subscribing to a channel is the most important consideration in deciding whether to sell rooms through it.					
7	The way that reservation transactions are delivered to the hotel is not important.					
8	It is essential that rate or inventory controls set by the hotel are correctly reflected through the online channel.					
9	Potential distribution partners must demonstrate their ability to deliver a large volume of business.					
10	Commission rates are basically the same for all electronic channels.					
11	My decision to allow rooms to be sold through a channel is not usually influenced by the company that owns the channel.					
12	Before I will allow rooms to be sold through a website, I must be assured that potential customers' trust and privacy are adequately protected.					
13	I always consider the brand image of a website, to decide whether or not it's suitable for my hotel.					
14	I pay particularly close attention to all the transaction costs (eg GDS fees, switch fees etc) associated with each channel.					
15	It is important to me that websites sites through which my rooms are sold achieve acceptable standards of speed and reliability.					
16	The target market of the website is not really important, as long as it produces bookings.					
17	I always try to avoid selling through websites that deliver reservations to the hotel by fax or telephone.					
18	I must have control over the prices that are displayed on partner sites.					
19	With all channel partners, I usually find out who owns the company and research them to see if they look like a viable business.					
20	I am most likely to choose to distribute my rooms through websites that offer the best customer experience.					

An Initial Study on the Intensifying Enterprises Information Management

Peiji Shao¹, Jingjing Li¹, Jing Yang¹, Jie Wan²

¹ School of Management, University of Electronic Science and Technology of China,
Chengdu 610054, China

² Johnson & Johnson, Philadelphia, PA, USA, 19104 USA

ABSTRACT

The information management has become an important base of developing an enterprise's eBusiness. Although our country has spent abundant manpower, material resources and financial support on enterprise information management, most of the information systems were not so good to meet our need and they didn't bring the effect into play sufficiently. Based on the two large questionnaires, this paper has made an initial study and analysis on the status and actual questions of Chinese information management. And then this paper studies the current status of information management of some of the dominant enterprises in Sichuan Province. At last this paper presents some suggestions of intensifying enterprises information management.

Keywords : eBusiness, information management, information resource management

In recent years, Chinese enterprises have spent abundant resources on enterprise informationization, but most of the information systems were not so good to meet our need and they didn't bring the effect into play sufficiently. So the benefit of enterprise information management is rather low. The author believes that one of the main reasons is short of 'intensifying enterprises information management'.

The development of enterprises information management in Sichuan Province is behind of the east. The ulterior study of enterprises information management in Sichuan Province is very important to promote the development of west.

1 THE ANALYSIS ON THE STATUS OF CHINESE INFORMATION MANAGEMENT

1.1 THE STUDY FOUNDATION

This paper has made an initial analysis on the status of Chinese information management that is based on the two large questionnaires.

The report 1^[1]: In 2001, the Economic Information Center of the State Economic & Trade Commission had made a questionnaire on the informationalization status, which reclaimed 638 effectual questionnaires. According to these questionnaires, they got valuable investigate result and analysis report. Among the investigated enterprises there are 231 national key enterprises and 510 state and state holding companies. The investigated enterprises cover the whole country and involve 33 industries such as mechanism, chemical industry, spin, electron, coal, automobile and so on, among which, state and state holding companies account for 79.94%, the other 12.23%, and oversize enterprises 15.36%, lager enterprise 59.4%, medium-sized enterprises 16.46%, other enterprise 2.66%.

The report 2^[2]: In 2002, the State Enterprise Information Work Leading Group Office and the Information Center of the State Economic & Trade Commission made the 4th questionnaire of the level of informationization construct and application of 520 state key enterprises, 120 experimental enterprise groups and district key enterprises, which reclaimed 570 effectual questionnaires, and according to these questionnaires they gained the investigate report of national enterprise informationization.

The two reports carry preferable authority and reflect the information management status of the dominant enterprises in China by some degree. As there are some difference between the reports, this paper integrate them and give a complete analysis of the information management status of the enterprises in China.

1.2 RESEARCH CONTENT

(1) The management status of enterprise information human resource

The investigation of information human resource all takes the first place in the twice questionnaire reports, which reflect that the persons with ability is very important in enterprise information management.

According to the second report, some of the investigated enterprises in 2002 pay more attention to the information management.

83.3% enterprises have set up the position of CIO which is corresponds to vice general manager; 89.1% have set up the position of full time personnel to run and maintain information system; 86.7% enterprises have rebuild and upgrade their system according to the development of technology and operation requirement; 74.9% enterprises have taken training informationization talent as a focus of their work; the proportion of setting up the position of CIO in 2002 have increase 14% compares to that in 2001. But the information skilled personnel just account for

0.72% of the employee total number, namely, there is lower than one information skilled person among every 100 employees. The compound backbone personnel who not only knows well IT ,but also is familiar with operation flow and enterprise management is quite lacking, which infect the development of enterprise information management.

(2) The foundation establishment construction of enterprise information management (the report 2)

The rate of possession of computers and access Internet is elevated. Every one hundred persons have 7 computers, every manager has 0.77 computer. In some enterprises, one person has computer. The number of computers, which access to Internet is about 68.4%.

Most enterprises have established intranet、internet and Portal. The coverage and application range of all kinds network are extended which have held every administrative levels, especially management department and core operations.

The method to access Internet of most enterprises has been close to advanced lever and the bandwidth is widened. The intranet bandwidth of 77.7% enterprises is 100 MB to 1000MB, 76.4% enterprises' desktop bandwidth come to 10-100MB and most enterprises' internet bandwidth is more than 1MB, which reaches the lever of high speed transmit channels.

Most enterprise have spent abundant fund and are provided with foundation. Many enterprises have build their portal, but the content isn't updated in time, and the columns of many enterprise portal are the same and lack of individuality, which is hard to bring benefit to enterprises.

(3) The construction of enterprise information system (EntIS, the report 1)

Tab.1 is the summarization of the Report 1.

The implementation rate of some information system is quite low, the rate of CAD/CAM is tiptop but just close to 9%.

Tab.1 The input and construct status of EntIS (unit: Million Yuan)

CAD/CAM	Total input	Construct status			
	12.97	Full realization	Partly realization	Not take measures	Unknown
		9%	50%	19%	22%
Basal MIS (OA, MIS)	19.42	8%	59%	18%	15%
Integrated MIS (ERP, CMIS)	0.86	5%	31%	44%	20%
eBusiness system	2.78	1%	23%	55%	21%

In 2000, there are just 138 enterprises putting eBusiness in practice, which is 21.6% and those of full realization eBusiness system only occupy 1%. The collectively Level of eBusiness is just the age of primary stage.

(4) the status of information share (the report 2)

The proportion of inter-linkage and mutual communication in different system is lower than one third most of which just is limited to the range of department and local computers. The number of full realization information share is 41, which just account for 7.2%, the number of partly realization information share is 336 which just is 59%, and the number of unfulfilment information share is 193 which is 33.9%.

Unfulfilment enterprise information share will cause Information isolated Island, and the superiority of information management can't be brought into play. The most important reason is lack of collective layout、system design and durative optimize strategy.

(5) The value brought about by using MIS (the report 2)

About 36.7% companies enhance the ability of deliver the goods in time; 77.5% reduce the cost(such as buying cost and overtime cost); 67% reduce the total production maintenance and productive time of a working; 64.6% increase the sales income; 66.3% increase the velocity of flexible capital. Information management increases the value and benefit of the company, the application focuses on the operation management, and few involve management and strategy level.

Comparing report 1 with report 2, we have the finds:

(1) In report 2, there is the investigation about collective layout of enterprise informationization which isn't mentioned in report 1 and enterprises pay more attention to collective layout of informationization.

(2) The key system content has been changed and enterprises pay more attention to the technology of Data Integration and Data Warehouse.

ò The systems of ERP、Intranet、EC、OA were made as the four key systems in the report of 2001.

ò According to the Statistics of report 2, the systems of ERP、OA、Intranet、data integration and data warehouse were the emphasis of next one-step information construction.

1.3 EXISTENT PROBLEMS

The information management has made some progress in china, but now the level is low. The main problems are:

(1) The focus of information management is in the data processing in a low level.

(2) Enterprises are lack of reasonable and effective collective layout in implement of information management.

(3) Enterprises' information resource has not been made good use and the lever of share information resource is low.

(4) They just can be used as to deal with some simple affairs and can't provide effective decision making to middle and high level.

2 THE CONTRAST ANALYSE OF THE INFORMATION MANAGEMENT IN SICHUAN'S PREPONDERANT ENTERPRISE

According to the Report 2, there are 360,000 industry corporations in Sichuan province. Among them, there are 292 large enterprises, 578 medium-sized enterprises, 4230 state-owned enterprises, 60,000 collectively-run enterprises, 290000 individual enterprise in town including other 3195 industry enterprise such as private enterprise, foreign investor and Hongkong, Macao, Taiwan investment enterprise and 73 stock limited companies. There are 5 well-known enterprises in Sichuan province, they are Changhong Co. Ltd., Lu Zhou Ancient Cellar, Wuliangye Group, Langjiu Group and Sichuan Jiannanchun Group Co. Ltd. Besides, there are 216 province's famous-brand enterprise and 230 province's famous-brand products.^[3]

2.1 RESEARCH SAMPLES

This paper choose 12 typical enterprises, they are Sichuan Changhong Electric Co. Ltd., Sichuan Electric Power Company, Pan Steel Group Co, Wuliangye Group Co. Ltd., Chengdu Aircraft Industry Corporation, Sichuan Jiannanchun Group Co. Ltd., Chengdu Enwei Group, Telecommunication Corporation of Sichuan province, Chengdu Cigarette Factory, Chengdu Pepple's Department Store (group) Co. Ltd., Dong Fang Steam Turbine Works, New Hope Group.^[4]

Tab. 2 is the rank of the top five enterprise in Sichuan^[5]

Tab. 2: the rank of the top five enterprise in Sichuan

Enterprise Name	Rank by sales	Rank by general asset	Rank by profit	Rank by tax
Changhong Co. Ltd.	1	5	3	6
Electric Power Company	3	1	9	4
Pan Steel Group Co.	2	3	2	3
Wuliangye Group Co. Ltd.	4	6	1	2
Chengdu Aircraft Industrt Corporation	11	10	96	157

2.2 DATAANALYSIS

Through analyzing the data of enterprise information management above-mentioned, there is a conclusion of the 12 enterprises information management.

(1) The status of department's MIS and the construction of LAN.

Tab.3 is the status of department's information management

Tab.3 The status of department's IM

	Already established	unknown
Realizing department's information management system	7	5
Intranet been established	9	3

Among the 12 enterprises selected, a majority of them have equipped with computers in their departments and established the intranet, the cover and application range of which have increased and expanded to every lever of enterprise, especially management department and core operations.

(2) Ways of information exchange between enterprises and outside

Table 4 is the Way of information exchange between the enterprise and outside

Mostly, those who have constructed websites are successful enterprises. Among them, the website construction of the Chengdu Enwei Group is a successful one. By the year of 2003, its website has provided more than 30 channels, 500 items and 2000^[6] web pages. Also, Sichuan Jiannanchun Group Co. Ltd. has set functions of inviting public bidding for providers and booking service on line, which ranges at the top of the website-construction.^[7]

Tab. 4 Ways of information exchange between the enterprise and outside

	Website	E-mail	Center of customer service
Number of enterprises	10	8	2

Among the 12 enterprises, there is 8 ones which have opened their own email boxes. Sichuan Electric Company has communicated interior information through email. Meanwhile, Pan Steel Group Co uses email for information transmission and has established the internet node web which can transmit files, gather and publish information by means of FTP、WWW etc.^[8] Likewise, Chengdu Aircraft Industry Corporation has realized the communication with some factories through internet, such as Boeing. Similarly, Changhong Co. Ltd. has

constructed the program of CIQ Computer-connected Supervision on machining and trading, and with the ERP system of SAP the efficiency of passing CIQ examination has been improved greatly^[9].

However, there are some problems existed as follows:

- ò Website columns are similar and there is lack of contents with characteristics.
- ò The supporting of enterprise's internet information is weak. The weak base of interior information management leads to that it is not timely and efficient enough to integrate, gather, transmit and share all kinds of information.

(3) Establishment of the information system

Tab.5 is the establishment of the Information System

Tab.5 Establishment of the Information System

	Already realized	Not yet	Unknown
CIMS	6	5	1
OA	7		5
ERP	2	10	0

As we can see from the table above, most of the manufacture enterprises have realized CIMS and OA is a major project of enterprise-construction. Basically, nearly every enterprise has realized the Financial MIS. And some enterprises have realized ERP system such as Changhong Co. Ltd. etc.

Among the 12 enterprises, the investment of information construction of Changhong Co. Ltd. has reached 100 million Yuan, and another 100 million Yuan will continue to be used to perfect the construction^[10]. Similarly, to popularize the enterprise's computer management, Chengdu Enwei Group has put the application at a higher level in the beginning. Several years ago, with large amount of money, Chengdu Enwei Group bought the PC system, top grade data storage and stock system, Sybase large relational database system from SUN Company. Meanwhile, other enterprises also have devoted much money into the information construction, such as Sichuan Electric Power Company, Sichuan Telecom, Chengdu Cigarette Factory, Pan Steel Group Co etc. As their conscious of information strengthened, the degree of information investment has increased.

However, there are still some problems existed: the degree of information construction fund devotion is still not enough. Take Changhong Co. Ltd for example, its total capital is 17637511700 yuan^[11]. But the absolute devotion proportion in information is merely 0.58%, which is far away from oversea enterprises normally about 8% to 10%.

(4) Application of information

Sichuan Jiannanchun Group Co. Ltd. has done quite well in gathering and using of information resources. And document information center of it plays double functions in document information storage, using and document administrative management and supervision. In the year of 2002, the corporation established the enterprise information base from its document library. With the international internet information from Xinhua News Agency, the corporation has directly saved more than 30million Yuan from material-purchasing, technology improvement, environment protecting and so such programs.

(5) Implement of eBusiness

Many enterprises have made primary attempts in eBusiness such as Chengdu Pepple's Department Store (group) Co. Ltd. , Chengdu Enwei Group , Sichuan Jiannanchun Group Co. Ltd.. But due to the limitation of capital, talent, the base of information, balance and safety of trading, the development of eBusiness still stays at a primary period.

2.3 Analysis of the results

As we can see from the above 12 dominant enterprises in Sichuan Province, their information management level stands for the advanced level of Sichuan Province in enterprise information management. However, these advanced enterprises are still at a lower level in information management. Thus, intensifying information management becomes an important task. Then from the present state of the enterprise information management in Sichuan Province, we can come to the following analysis:

(1) Status of basic facilities of enterprise information management

- ò Hardware. Computer application has been popular in most enterprises of Sichuan and they can deal with data with computers.
- ò Network construction. According to report 1, in the year of 2000, among the key enterprises in Sichuan, 54% of them have established their websites, and 13% have simple web pages, but still 29% haven't established yet. By the year of 2002, the proportions of enterprises that have established websites have increased to 71%. Many famous enterprises such as Changhong Co. Ltd and Wuliangye Group Co. Ltd. own their special websites and domain name.

(2) Construction of information system

Part of those enterprises, we take DongFang Steam Turbine Works for example. According to its trait as a discrete manufacture enterprise, it has realized enterprise information management, which helps the corporation to have been a supplier of America GE Company. Another part of those enterprises such as Chengdu Xuguang Electronics Co. Ltd., according to its characteristic in enterprise production and management, has designed and explored application software. Then step by step, it has realized information integration and data sharing of the process of purchasing, sale, stock, production and finance. Moreover, the functions of supervising, controlling, coordination become more normative, which strengthen the competition ability of the enterprise.

(3) Management software

Most enterprises adopt the homemade software, which are suitable for the enterprises' actuality, such as Chengdu Cigarette Factory. The software can basically satisfy the present need of the enterprises, with better services and the cost is much lower than foreign software.

The main problems existing in enterprise information management of Sichuan Province are as follows: Low benefit of information management; lacking of efficient total layout; still focusing on low level data processing; low level of information resource sharing; lacking of efficient strategy support; weak economic strength as a whole of Sichuan enterprises, even smaller amount of money can be used in information management; serious flowing-out of informative people in Sichuan enterprises. Thus, intensifying information management of Sichuan enterprises should be paid special attention to, through which enterprises of Sichuan can strengthen their key competition ability efficiently.

3. STRATEGY OF INTENSIFYING INFORMATION MANAGEMENT OF SICHUAN ENTERPRISES

The process of intensifying information management in enterprises is complex. Then we can consider the following aspect as a manager.

3.1 MAKING STRATEGIC PLANNING ON ENTERPRISE INFORMATION RESOURCE MANAGEMENT (IRM)

Making strategic layout on enterprise IRM has a quite important meaning in intensifying enterprise information management. Fig. 1 shows the process.

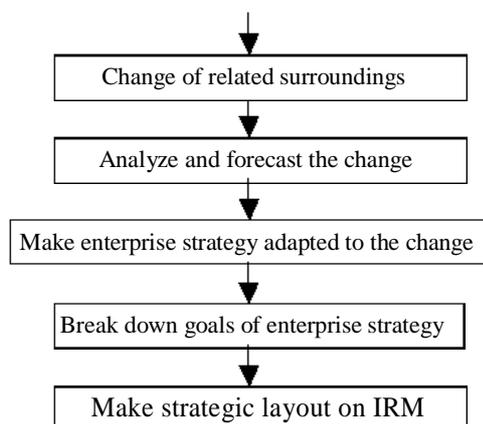


Fig. 1 Process of whole layout

3.2 STRENGTHENING THE ENTERPRISE' IRM

To strengthen enterprise' IRM, we can consider the following aspects: to make clear of the main task of enterprise IRM; to make strategic layout of enterprise IRM; to cultivate the information conscious of personnel; to set up right information concept; to set basic standard of IR; to improve the status of information management department and set related data management persons; to make plans that are scientific and reasonable in IR distribution and using; to develop and make use of information reasonably and efficiently; to supply different information according to different needs.

3.3 CONSTRUCTING ERP

It is important of BPR to intensify enterprise information management. The idea of BPR was first raised by Hamer and Qianpin^[13]. Literature [14] puts up the idea of process management, which centralizes on constructing end to end business process in standardization, and it is a systematical method with the aim of continuing to improve the business performance of the organization. The idea of process management calls for systematization and continuous improvement of process, so as to reduce the manual resistance in practice, which is more suitable for Chinese enterprises.

Process management reduces the risk of BPR. The average level of the enterprise development in China is fall behind than developed countries. Therefore, in the process of putting BPR into practice, a complete, thorough process will bring up great risk to enterprises. Thus, of enterprises' managers should consider carefully before using the method of BPR, according to their ability to stand risks.

4. CONCLUSIONS

This paper has analyzed two authoritative sample survey materials of enterprise information in China. Besides, there are also some researches and analysis on the examples of 12 typed enterprises information

management in Sichuan. From these analyses, we find there are some common problems such as the information system aren't in full use and the benefit of information management is in a low state. One important reason of these problems is that enterprises didn't pay high attention to intensify information layout on enterprise IRM. Another is to strengthen the enterprise IRM. And the last one is to construct the enterprise MIS on the base of BPR.

REFERENCES

- [1] The report of the status of Chinese enterprise informationization construction and the '10th Five Year Plan' [EB/OL]. <http://www.chinabbc.com.cn/yljc/view.asp?newsid=20041211167835&classid=>, 2002-2-10
- [2] The investigation of enterprise informationization in 2002 [EB/OL]. <http://www.setc.gov.cn/qyxxhjyxal/200302130071.htm>, 2003-02-13
- [3] General situation of enterprises in Sichuan province [EB/OL]. <http://www.sc.gov.cn/html/qiyezhaixian/CQGK.asp?id=13>, 2003-04-06
- [4] The list of the dominant enterprises in Sichuan Province [EB/OL]. <http://www.ctisc.com.cn/download/scys.htm>, 2003-04-15
- [5] The rank of 500 industry enterprises in Sichuan Province [EB/OL]. <http://www.scjm.gov.cn/gyqypx500asp?id=15>, 2002-03-18
- [6] The prospect of Chengdu Enwei Group [EB/OL]. <http://www.enwei.com.cn/medicine/bussiness/dsptext.asp?file=20029240203050001.htm&lmdm=020305>, 2003-04-16
- [7] The purchase of material [EB/OL]. <http://www.jnc.com.cn/yclcg/index.htm>, 2003-04-03
- [8] The informationization construction of Pan Steel Group Co. [EB/OL]. <http://www.scmie.com.cn/msmie/anli/pangang.asp>, 2003-05-12
- [9] The experience of informationization implement of Sichuan Changhong Electric Co. Ltd. [EB/OL]. <http://www.cseek.com/Enterprise/218718151532085248/20040713/1830464.shtml>, 2004-07-13
- [10] The informationization construction of Sichuan Changhong Electric Co. Ltd [EB/OL]. http://chengdu.ccw.com.cn/news/200209/0916_01.asp, 2002-09-16
- [11] The bulletin of public listed company in Shanghai Market [EB/OL]. <http://www.eefoo.com/xxzycontent/2002/0328/200203283003.asp>, 2002-03-27
- [12] Raymond McLeod, George Schel Management information systems—Theory and practice in the guide of management (translated by Zhang Chenghong) Beijing: Publishing House of Electronics Industry, pp27-29, 2002
- [13] Wu Xiaoming. BPR: Business Process Reengineering. *The transaction of Institute of Chongqing Industry Management*, Vol. 13, No. 1, pp44-48, 1999
- [14] Huang Aizhou, Mei Shaozu. Surpass BPR — Research on Management Thought of Process Management. *Scientific and Technology Management*, No. 12, pp105-107, 2002

Antecedents of E-Business Assimilation in Manufacturing SMEs

Louis Raymond¹, François Bergeron², Sam Blili³

¹ Département des sciences de la gestion, Université du Québec à Trois-Rivières, P.O. Box 500, Trois-Rivières, Québec, Canada, G9A 5H7 (louis_raymond@uqtr.ca)

² Groupe RGTI, Faculté des sciences de l'administration, Université Laval, Québec, Québec, Canada, G1K 7P4 (francois.bergeron@fsa.ulaval.ca)

³ Institut de l'Entreprise, Université de Neuchâtel, CH-2000 Neuchâtel, Switzerland (sam.blili@unine.ch)

ABSTRACT

In order to further our knowledge on the use of the Internet and the Web in manufacturing SMEs, the present research seeks through an empirical study of 108 Canadian firms to explore the following questions: For what purposes are the Internet and the Web presently used, i.e., to what extent are e-business functions assimilated in manufacturing SMEs? What characteristics of the SMEs' environmental, strategic, managerial, operational, and technological context are associated with e-business assimilation?

Keywords: e-business assimilation, Internet use, Web use, SME, small business, manufacturing, growth, export, performance.

1. INTRODUCTION

Many small and medium-sized enterprises (SMEs) presently evolve in a complex business environment, characterized by globalization, the internationalization of markets, and the need for greater efficiency, effectiveness, and competitiveness based on innovation and knowledge. This has put increasing pressure upon the management of these firms, especially manufacturing SMEs that must now compete globally [1]. In order to lower their operating costs, increase productivity and quality, and respond to the increased requirements of their customers and other business partners, a number of these firms scan the technological environment and make sizable investments in adopting Internet-based or World Wide Web-based technologies as the infrastructure for e-business applications [2]. To the extent that e-business is assimilated by the SME, it can significantly affect the firm's key business processes and relationships such as servicing customers and collaborating with business partners [3].

The adoption and assimilation of information technology, and of the Internet in particular, are deemed to be influenced by a number of environmental, organizational, technological, and individual factors [4,5]. While a significant amount of research has been done on the determinants of e-business in large firms, much less is known in regard to SMEs, especially manufacturing SMEs, given that their use of the Internet is usually more recent and potentially subjected to different constraints and contingencies related to their specificity as organizations. The specific characteristics that fundamentally distinguish SMEs from large enterprises relate to their environment, structure, strategy and decision process, and psycho-sociological context such as the dominant role of the owner-manager, but also relate to their flexibility, proximity to markets, and quickness to react and reorient themselves [6]. For instance, the fact that

smaller manufacturers often act in a subcontracting capacity or in partnership with other firms within network enterprises can influence their assimilation of e-business [7]. Operating most often in "make-to-order" or "job-shop" production environments, rather than selling products as such to their customer, these firms "sell" their manufacturing knowledge and know-how within a particular industry (say, plastics) or industrial activity (say, machining). This makes for a more complex commercial cycle linking customers, suppliers, and other business partners [9].

In order to further our knowledge on the use of the Internet and the Web in manufacturing SMEs, the present research seeks through an empirical study of 108 Canadian firms to explore the following questions: For what purposes are the Internet and the Web presently used, i.e., to what extent are e-business functions assimilated in manufacturing SMEs? What characteristics of the SMEs' environmental, strategic, managerial, operational, and technological context are associated with e-business assimilation?

2. RESEARCH MODEL AND HYPOTHESES

Founded on the Tornatsky and Fleisher's [9] technology-organization-environment (TOE) framework, the research model is presented in Figure 1. Five constructs that represent the firm's environmental, strategic, managerial, operational and technological context are deemed to influence a manufacturing SME's assimilation of electronic business activities. In this study, the organizational assimilation of e-business is conceptualized as the use of the Internet (including intranet and extranets) and the Web. The first business aim of this use is communicational and informational (e-communication), the second is more strategic in nature (e-business intelligence) while the third is transactional and collaborative (e-commerce and e-collaboration). From this point of view, e-business assimilation is deemed to be more extensive when the

firm uses the Internet and the Web in support of more aims.

2.1 Environmental context of e-business in manufacturing SMEs

Many small manufacturers must now operate within a business environment characterized by the emergence of new organizational forms such as the network or extended enterprise, whose information infrastructure is based upon inter-organizational systems and technology [10]. In a networked business environment, manufacturing SMEs would require Internet and Web-based capabilities to better support supply chain management and customer relationship management in particular. The emphasis placed upon communication and cooperation in such an environment is conducive to the assimilation of the B2B form of e-business. Greater use of the Internet and the Web would thus result from increased networking intensity on the part of the SME, that is, from the establishment of more partnerships with its customers, distributors, suppliers, competitors and other business partners such as research centers and consultants.

H1 - The greater the firm's networking intensity, the greater its assimilation of e-business.

2.2 Strategic context of e-business in manufacturing SMEs

A firm's strategic orientation or posture is its response to its environment. As this environment becomes more hostile or complex, SMEs with a more aggressive or more entrepreneurial orientation increase their competitiveness by seeking new markets and putting emphasis on technological leadership and new product/market combinations [11]. Increased complexity in the environment is also seen as causing the acquisition of competitive, market and technological information to be more continuous, variant, and wide-ranging [6]. The strategic orientation of the SME must also be in alignment with its strategic management of information technology [12], including the Internet [13]. Given the goals of a more aggressive strategy in terms of market, product and technological developments, one expects e-business to be called upon more in support of such developments.

H2 - The more aggressive the firm's strategic orientation, the greater its assimilation of e-business.

2.3 Managerial context of e-business in manufacturing SMEs

Small firms are deemed to be "organic" to the extent that their strategy, structure, and culture are embodied by their owner-managers. The primary goals and characteristics of entrepreneurs are thus crucial in determining the firm's innovativeness and orientation toward product/market development and technological sophistication [14]. In this regard, small business studies have shown that the previously acquired knowledge and experience of business owner-managers condition their

behavior as adopters of information technology [15]. In addition, a key component in the small firm's e-business learning experience is the owner-manager's individual learning [10]. Domain-specific knowledge that comes with experience in a specific industry or sector as well as more general knowledge obtained from a higher education would thus influence the entrepreneur's awareness of the various e-business functions and applications to be assimilated by the organization, hence the third research hypothesis:

H3 - The greater the firm's owner-manager experience and education, the greater its assimilation of e-business.

2.4 Operational context of e-business in manufacturing SMEs

The appropriate choice of advanced manufacturing technologies is influenced by the manufacturing process structures used by an organization, i.e., by its production environment [16]. The flexibility required within a "make-to-order" production environment has led SMEs to adopt computer-integrated manufacturing [17]. In parallel fashion, make-to-order companies are seen to increase their level of vertical integration and improve their manufacturing planning and control decisions through e-business applications that allow them to better communicate and exchange information with their customers and suppliers [18]. Given the greater complexity of buyer-seller relationships in a make-to-order production environment, the fourth hypothesis follows:

H4 - The more the firm's production environment is of the "make-to-order" type, the greater its assimilation of e-business.

2.5 Technological context of e-business in manufacturing SMEs

Advanced manufacturing technologies (AMT) are deemed to "significantly impact the design and outcomes of core organizational processes" [19, p. 995]. In looking at the links between strategy, AMT and performance, some studies have stressed the information processing capability inherent in AMT in addition to the flexibility dimension [20]. By linking with the information outputs and inputs of internal systems and applications such as CAD/CAM, MRP II and ERP, e-business applications can increase the manufacturing SME's level of external integration through their enabling of design and production information exchanges with business partners, coordination of trans-organizational business processes (e.g., just-in-time or synchronous production with "extended" ERP), and inter-organizational collaboration within a networked environment [3]. Given the compatible and complementary nature of e-business and advanced manufacturing applications, the fifth hypothesis follows:

H5 - The greater the firm's assimilation of advanced manufacturing technologies, the greater its assimilation of e-business.

3. METHODOLOGY

The data used in the study were obtained from the PDG[®] database [21], containing information on more than 350 manufacturing SMEs located in Quebec, Canada. Out of these, 108 completed a questionnaire about their use of the Internet and the Web. The number of employees of the sampled organizations ranges from 19 to 336, with a median of 60. Over fifteen industrial sectors are present, including metal products, wood, plastics and rubber, electrical products, food and beverages, and machinery.

In assessing the environmental context, the firm's networking intensity is measured by the number of distribution, marketing, design and R&D partnerships established with customers, suppliers and other third-parties such as research centres [22]. As the key informant on the firm's strategic orientation, the CEO is asked to rate the aggressiveness with which new markets were developed and new products are launched, and the proactiveness with which new technologies are introduced, on scales adapted from Covin and Slevin [23] and Julien and Raymond [24]. The production environment is assessed by the proportion of total production that is done in "make-to-order" mode as opposed to mass or continuous production modes [25]. Brandyberry *et al.*'s [19] classification is used to assess the firm's level of AMT assimilation: level 1) stand-alone AMT, level 2) functionally oriented AMT (CAM and FMS), and level 3) computer-integrated manufacturing. Growth is assessed by the average sales increase over the last three years whereas internationalization is measured by the mean percentage of sales turnover that has been generated via exports over the last two years.

The measure of e-business assimilation is based on a list of business activities for which the Internet and the Web is used by the firm. The grouping of these various functions corresponds to various stages of e-business assimilation, as adapted to the manufacturing SME context from previous studies [26,27,28,29]. The extensiveness of each type of use is measured by counting the number of activities checked by the respondents.

4. RESULTS

The partial-least-squares (PLS) method was chosen to test the hypotheses. The criteria of unidimensionality, reliability and discriminant validity of the measures were considered to be met. The research hypotheses are tested by examining the direction, strength and level of significance of the path coefficients calculated by the PLS method, as shown in Figure 2.

H1: The path coefficients first confirm that the environmental context influences the extent to which the sampled manufacturing SMEs have assimilated electronic business. The firm's networking intensity,

both upstream and downstream in the value-chain in the form of design, R&D, and marketing partnerships, is found to determine their use of the Internet and the Web for transaction or collaboration and business intelligence purposes, but not for communication or information purposes.

H2: Manufacturing SMEs whose strategic orientation is more aggressive in regards to new markets, products and technologies are found to assimilate e-business more extensively for purposes of communication or information, and transaction or collaboration with business partners, but not for business intelligence purposes.

H3: The hypothesis that the managerial context influences the assimilation of e-business is partly confirmed. SMEs are found to use e-communication, e-commerce and e-collaboration more when led by a chief executive with more experience in the sector of activity.

H4: A significant but negative path coefficient associates the production environment with the e-commerce and collaboration dimension of e-business. This result would indicate that the greater complexity of the inter-organizational business processes associated with "make-to-order" production make them more difficult to reengineer with Internet and Web-based technology, as opposed to the simpler catalog-based buying and selling processes associated with "make-to-stock" production.

H5: The technological context was partly confirmed as a determinant of e-business assimilation in small manufacturing firms. More precisely, the path coefficients indicate that the more integrated are SMEs in terms of their assimilation of advanced manufacturing technologies, the more they use the Internet and the Web for business intelligence purposes.

Globally, these results indicate that the hypothesized antecedents of e-business assimilation in manufacturing SMEs explain a significant percentage of variance in e-communication (20 %), e-business intelligence (12 %), and e-commerce and collaboration (23 %). In terms of growth, the assimilation of e-business by manufacturing SMEs explains 4 % of these firms' sales growth and 5 % of their export performance as there are obviously many other factors that contribute to these firms' performance.

5. CONCLUSION

With the advent of global competition and new organizational forms based on networks of cooperating firms, the successful assimilation of e-business is bound to take added importance for many SMEs in terms of survival, growth, and competitiveness. Given the dearth

Figure 1: Antecedents of e-business assimilation in manufacturing SMEs

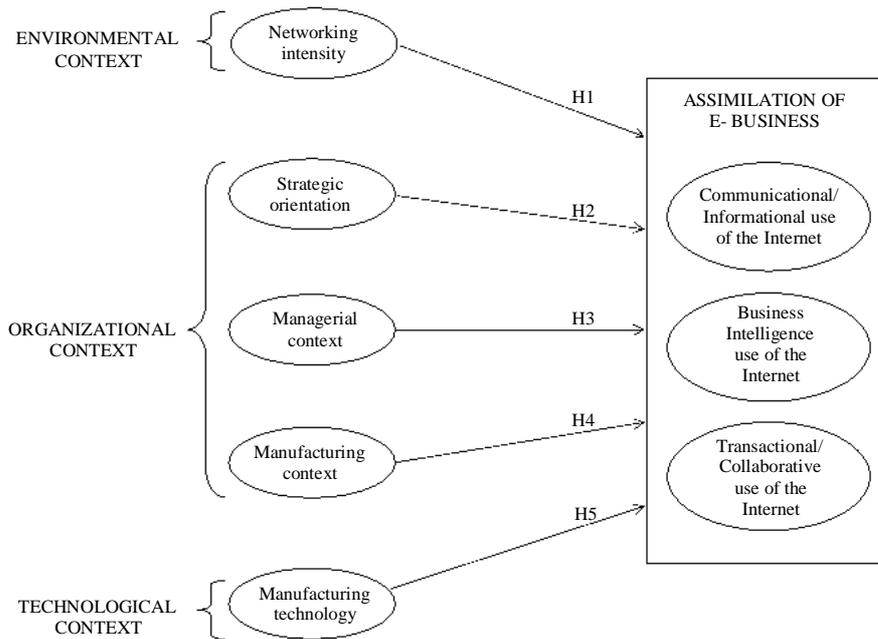
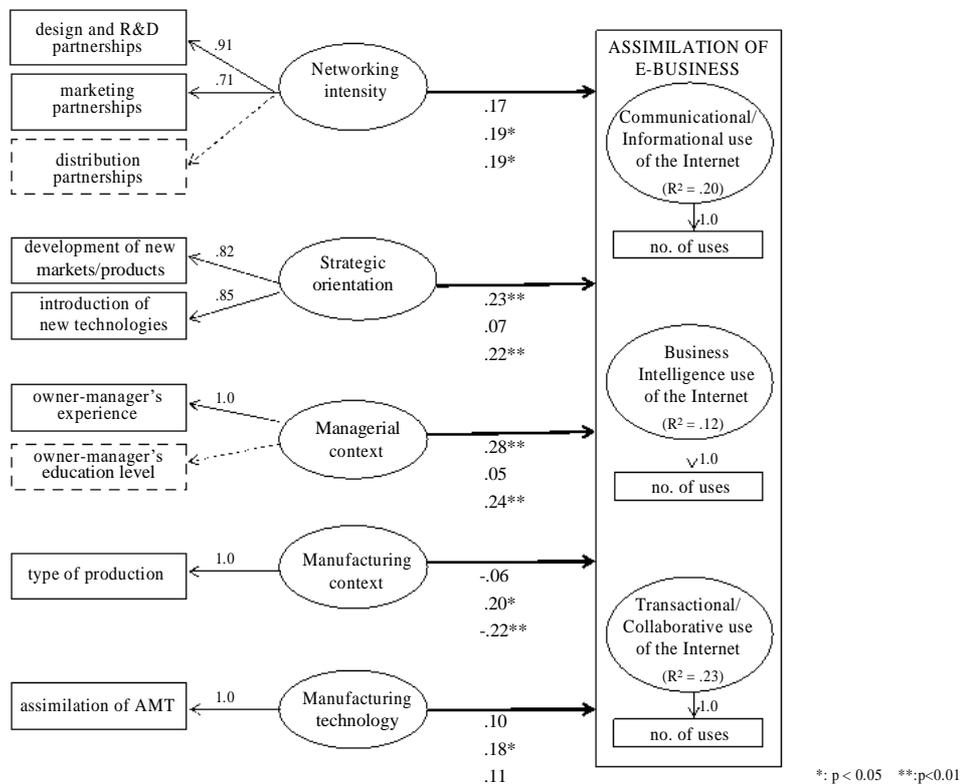


Figure 2: Results of the structural model analysis (PLS, n = 108)



of empirical knowledge in this regard, the present study has contributed to a further understanding of the nature and state of e-business assimilation in small manufacturing firms, and of the antecedents and market development outcomes of this assimilation. It

is recognized that these organizations are highly flexible and adaptable to change, be it environmental, organizational, or technological. Some of these already possess advanced manufacturing systems and, in the now global business environment, must follow suit

with Internet and Web-based information systems such as e-business intelligence, e-commerce, and e-collaboration systems to improve their competitive position. E-business investments cannot insure greater performance unless they are coherent with the competitive environment, strategic goals, and business processes of manufacturing SMEs. Results of this study imply that future research on the assimilation of e-business in these organizations must further explore the competitive dimensions of this phenomenon. In this regard, better explanations depend on furthering our knowledge of the level of alignment or “fit” between the external (environmental uncertainties and pressures, networks) and internal (entrepreneur, strategy, structure, technology, culture) contexts in which e-business applications are introduced.

ACKNOWLEDGEMENTS

The authors would like to thank the Canada Research Chairs Program and the Canada Foundation for Innovation for their financial assistance.

REFERENCES

- [1] Cagliano, R.K. and Spina, G. (2002) “A Comparison of practice-performance models between small manufacturers and subcontractors” *International Journal of Operations & Product Management* 22 (12): 1367-1388 [17]
- [2] Kleindl, B. (2000) 'Competitive Dynamics and the New Business Models for SMEs in the Virtual Marketplace' *Journal of Developmental Entrepreneurship* 5 (1): 73-85.
- [3] Norris, G., Dunleavy, J.R., Hurley, J.R., Balls, J.D. and Hartley, K.M. (2000) *E-business and ERP: Transforming the Enterprise*, New York: John Wiley & Sons.
- [4] Armstrong, C.P. and Sambamurthy, V. (1999) 'Information Technology Assimilation in Firms: The Influence of Senior Leadership and IT Infrastructures' *Information Systems Research* 10 (4): 304-327.
- [5] Dholakia, R.R. and Kshetri, N. (2004) 'Factors Impacting the Adoption of the Internet among SMEs' *Small Business Economics* 23 (3): 311-322
- [6] Julien, P.-A., Raymond, L., Jacob, R., and Ramangalahy, C. (1996) 'Patterns and Determinants of Technological Scanning: An Empirical Investigation of Manufacturing SMEs' in Reynolds, P.D., S. Birley, J.E. Butler, W.D. Bygrave, P. Davidsson, W.B. Gartner and P.P. McDougall (eds.), *Frontiers of Entrepreneurship Research*, Babson Park, Massachusetts: Babson College, 584-598.
- [7] Raymond, L. and Blili, S. (1997) 'Adopting EDI in a Network Organization: The Case of Subcontracting SMEs' *European Journal of Purchasing and Supply Management* 3 (3): 165-175.
- [8] Cagliano, R., Blackmon, K. and Voss, C. (2001) 'Small Firms Under MICROSCOPE: International Differences in Production/Operations Management Practices and Performance' *Integrated Manufacturing Systems*, 12 (7): 469-482.
- [9] Tornatsky, L.G. and Fleischer, M. (1990) *The Processes of Technological Innovation*, Lexington, Massachusetts: Lexington Books.
- [10] Raymond, L. and Blili, S. (2000) 'Organizational Learning as a Foundation of Electronic Commerce in the Network Organization' *International Journal of Electronic Commerce* 5 (2): 29-45.
- [11] Özsoymer, A., Calantone, R.J. and Di Benedetto, A. (1997) 'What makes firms more innovative? A look at organizational and environmental factors' *Journal of Business & Industrial Marketing* 12 (6): 400-416.
- [12] Bergeron, F., Raymond, L. and Rivard, S. (2001) 'Fit in Strategic Information Technology Management Research: An Empirical Comparison of Perspectives' *Omega* 29: 125-142.
- [13] Porter, M.E. (2001) 'Strategy and the Internet' *Harvard Business Review* 79 (3): 62-78.
- [14] Miller, D. (1993) 'The architecture of simplicity' *Academy of Management Review*, 18 (1): 116-138.
- [15] Riemenschneider, C.K. and Mykytyn, P.P. (2000) 'What Small Business Executives Have Learned About Managing Information Technology' *Information & Management*, 37: 257-269.
- [16] Kathuria, R. and Igarria, M. (1997) 'Linking IT Applications with Manufacturing Strategy: An Integrated Framework' *International Journal of Operations & Production Management* 17 (6): 611-629.
- [17] Ari Samadhi, T.M.A. and Hoang, K. (1995) 'Shared Computer-Integrated Manufacturing for Various Types of Production Environment' *International Journal of Operations & Production Management* 15 (5): 95-108.
- [18] Olhager, J. and Rudberg, M. (2003) 'Manufacturing Strategy and E-business: An Exploratory Study' *Integrated Manufacturing Systems* 14 (4): 334-345.
- [19] Brandyberry, A., Rai, A. and White, G.P. (1999) 'Intermediate Performance Impacts of Advanced Manufacturing Technology Systems: An Empirical Investigation' *Decision Sciences* 30 (4): 993-1020.
- [20] Kotha, S. and Swamidass, P.M. (2000) 'Strategy, Advanced Manufacturing Technology and Performance: Empirical Evidence from U.S. Manufacturing Firms' *Journal of Operations Management* 18 (3): 257-277.
- [21] St-Pierre, J. and Delisle, S. (2004) 'An Expert Diagnosis System for the Benchmarking of SMEs' Performance' *Benchmarking: An International Journal* (in press).
- [22] Gulati, R. (1998) 'Alliances and Networks' *Strategic Management Journal* 19: 293-317.
- [23] Covin, J.G. and Slevin, D.P. (1989) 'Strategic Management of Small Firms in Hostile and Benign Environments' *Strategic Management Journal* 10: 75-87.

- [24] Julien, P.-A. and Raymond, L. (1994) 'Factors of New Technology Adoption in the Retail Sector' *Entrepreneurship Theory and Practice* 18 (4): 79-90.
- [25] Mechling, G.W., Pierce, J.W. and Busbin, J.W. (1995) 'Exploiting AMT in Small Manufacturing Firms for Global Competitiveness' *International Journal of Operations & Production Management* 15 (2): 61-76.
- [26] Raymond, L. and Bergeron, F. (1996) 'EDI Success in Small and Medium-Sized Enterprises: A Field Study' *Journal of Organizational Computing and Electronic Commerce* 6 (2): 161-172.
- [27] Kula V. and Tatoglu, E. (2003) 'An Exploratory Study of Internet Adoption by SMEs in an Emerging Market Economy' *European Business review* 15 (5): 324-333.
- [28] Levy, M. and Powell, P. (2003) 'Exploring SME Internet Adoption: Towards a Contingent Model' *Electronic Markets* 13 (2): 173-181.
- [29] Xu, S., Zhu, K. and Gibbs, J. (2004) 'Global Technology, Local Adoption: A Cross-Country Investigation of Internet Adoption by Companies in the United States and China' *Electronic Markets* 14 (1): 13-24.

Business Drivers for Application Servicing and a Software-as-a-Service Model

Jiri Vorisek

University of Economics Prague, W. Churchill Sq. 4, 130 67 Prague, Czech Republic
vorisek@vse.cz

ABSTRACT

Many organizations are re-evaluating their approach to delivering enterprise applications, and are looking for ways to control IT costs. There is growing evidence of reluctance to fund large-scale implementation projects, and of tighter budgets forcing more careful cost-benefit analysis to justify IT investments. These efforts to control IT costs lead to consideration of alternative models for delivery of enterprise applications, such as the Software-as-a-service model or ASP (Application Service Providing).

This paper discusses the business drivers for application servicing and gives a detailed description of the Software-as-a-service model in comparison to traditional one. It brings the arguments that the Software-as-service model constitutes a viable alternative to software licensing for many application types today, and will become the dominant method for delivery of enterprise applications in not too distant future.

Keywords: Software-as-a-service model, ASP, application service providing, Utility Computing

1. INTRODUCTION

The traditional approach to the implementation and operation of enterprise applications typically involves the purchase of hardware and software platforms from technology vendors and their implementation within the organizations, in most cases drawing on extensive assistance from external consulting organizations, in order to tailor the application software systems to the specific needs of the organization.

This model has been used extensively in the context of ERP systems, and has been applied to other types of application including CRM, Data Warehousing, and Business Intelligence applications. Almost invariably, situations where organizations own and maintain their entire IT infrastructure and application systems, lead to very high cost of ownership, and consequently high levels of IT spending which can detract from the core business that the organization is engaged in. Gartner research estimated that big global banks allocated 15-20 percent of their operating budget to IT in 2002; this percentage had grown by 50-60% over five years. There is now increasing evidence that traditional model, which involves *ownership* of software and the associated IT infrastructure, is not sustainable in the long term. Significantly, many businesses are now doubtful about IT benefits altogether, and are no longer prepared to fund large-scale implementation projects [3]. Some even argue that the productivity improvements once assumed to be the result of technology investment are an aberration related to other factors such as longer working hours [16]. In an attempt to improve ROI (Return on Investment) and improve focus on core business many organizations have adopted partial, and in some cases full outsourcing of IT systems. However, companies that provide outsourcing services tend to use the traditional software implementation and maintenance approach,

and hence suffer from similar high levels of cost.

As a result of these developments IT vendors are intensively evaluating more cost-effective models for the delivery of enterprise applications. In late nineties, a new *application service model* has emerged as an alternative to the traditional approach, mainly in the form of providing application services over the Internet. Delivering applications as services has a number of advantages from the point of view of user organizations.

This paper discusses the business drivers for application servicing and gives a detailed description of the Software-as-a-service model in comparison to traditional one. It brings the arguments that the Software-as-service model constitutes a viable alternative to software licensing for many application types today, and will become the dominant method for delivery of enterprise applications in not too distant future.

2. BUSINESS DRIVERS FOR APPLICATION SERVICING

In this section we consider the main business drivers that are forcing the transition to a new model for the delivery of enterprise applications.

2.1 High cost of IT projects

Problems of controlling the costs associated with IT projects are well documented. Notwithstanding the long experience that the IT industry has with the implementation of enterprise applications, costs of many projects significantly exceed their original budgets. According to a study of ERP implementation projects of 117 US companies 25% exceeded their

budgets, 20% were abandoned before completion, and 40% failed to achieve business objectives [17]. There have been many similar studies of this type that clearly demonstrate that the traditional model of implementation and maintenance of enterprise applications is associated with significant risks that are not being addressed by improved implementation methodologies and more technologically advanced software platforms. Outsourcing of the implementation and maintenance of enterprise application does not always bring anticipated benefits as the implementation methodologies and technology architectures used by outsourcing organizations do not differ substantially from methodologies used by end-user organizations.

2.2 Fast rate of technology change

Another significant risk factor associated with enterprise applications is the rapid development of underlying technologies requiring costly and time-consuming updates. There is growing evidence that end-user organizations are unable to absorb new technologies at the rate that vendors produce them, and are locked into a *vicious* cycle of migrations and upgrades. New technology platforms, or new versions of existing applications are often forced on end user organizations by technology vendors who are reluctant to support older versions of their products; implementation of new versions may not lead to any direct business benefit.

2.3 High demand on IT skills

The traditional model of implementation of enterprise applications is associated with high demand on IT skills, but many small and medium size enterprises (SMEs) cannot afford to maintain their own internal IT staff. Another contributing factor is that the expertise of IT specialists employed by end-user organizations and third-party consulting companies is often not up-to-date and significantly lags behind the expertise available from the technology vendors. This leads to poor implementation results, and is a major cause of high-rate of failure of IT projects.

2.4 Complexity of ERP systems

ERP vendors have attempted to deliver a comprehensive enterprise application solution that satisfies the requirements of a large number of organizations irrespective of the industry and the needs of individual businesses. This approach resulted in highly complex solutions leading to increased cost of ERP solutions that the client organizations are no longer willing to accept. The utilization of the overall functionality of an ERP system by end users is relatively low. Customization to suit the needs of individual customers requires high level of expertise to setup a large number of configuration parameters. Lack of knowledge of client's business processes by the consultants is another

key factor according to recent studies that inhibits successful implementation of ERP systems [13]. ERP systems are characterized by high complexity of operation even in situations where the corresponding business process is relatively simple, and that leads to high cost of training of end users. The demands on hardware resources increase as new versions of the software are released.

2.5 Globalization of business environment

Globalization has two important influences on enterprise applications. Firstly, as a result of globalization and the formation of regional economic blocks with standardized business processes and regulations, ERP systems are becoming more generally applicable and require less customization to suit individual countries. This is particularly significant in the European Union with growing number of member countries using identical business laws and regulation; this will eventually lead to standardization of ERP applications across the entire region. Second important influence is the global deployment of enterprise applications. Many large companies with worldwide offices implement global applications across the entire enterprise, in some cases using a single centralized data center and centralized applications accessible via the company's Intranet. This simplified environment, with standardized business processes across the entire organization is highly suited to implementation using the ASP model.

2.6 Increased acceptance of outsourcing

There has been an increase in awareness of alternative models for the implementation of enterprise applications, in particular of the outsourcing model. Organizations are aware of the need to focus on their core business, and are prepared to consider outsourcing other business functions, including their ERP applications.

3. RECENT TRENDS IN APPLICATION SERVICING

Application Service Providing (ASP) emerged towards the end of 90s with claims of extensive advantages for client organizations, in particular for SMEs. Notwithstanding many perceived advantages the ASP approach has not gained wide acceptance as the new model for delivery of enterprise applications. Many of the early ASP providers have not been able to establish a viable business model for application servicing, and have discontinued ASP services, or went out of business altogether. Other factors contributing to the failure of early ASP providers included lack of a suitable technological infrastructure for hosting a large number of complex enterprise applications in a scalable and secure manner, poor customisation capabilities, and almost total lack of integration facilities. As a

result of these shortcomings, early ASP providers failed to deliver major cost savings to their customers, resulting in poor acceptance of application servicing by the market place.

Recently, however, a number of important IT vendors have re-confirmed their commitment to application servicing in the context of the new Utility Computing approach, and have made large investments in infrastructure for the delivery of application services [6]. Given earlier experiences with traditional outsourcing and ASP most user organizations remain sceptical about ASP and are waiting to see if the benefits are going to be realized as claimed by the vendors.

3.1 IT business factors

As a result of the recent IT downturn the sales of new licenses for enterprise application software have stagnated and in some cases declined. There is some evidence that as the enterprise application software market matures, major ERP vendors are changing their revenue model to decrease their reliance on new software licenses towards income generated from software license upgrades and product support [18], [19]. This combined with the fact that most organizations spend as much as 80% of software-related costs on software maintenance and related activities [11], creates a situation where licensed software is de-facto *rented*. *It is precisely this high-level of on-going costs that motivate many organizations towards alternatives such as outsourcing and application servicing.*

3.2 Emergence of Utility Computing

The main idea of Utility Computing is that IT services are supplied on demand (i.e. as required by the end-user organization) via a grid of interconnected, dynamically configurable, highly reliable and scalable computing resources (i.e. servers, storage, and applications). Computing grid provides an ideal infrastructure for application servicing as it can host a large number of ASP applications in a scalable and reliable manner. *Resource sharing and improved hardware utilization of grid computing environments provides a more cost effective solution for hosting enterprise applications than a set of independent servers each dedicated to a specific application.*

A number of infrastructure vendors (IBM, HP, Oracle) are in the process of building large data centres with the view of moving towards the Utility Computing model [7]. *Investment in infrastructure on this scale clearly demonstrates a strategic commitment to Utility Computing and more specifically to application servicing as the new outsourcing model for enterprise applications.* Recent efforts to standardize Utility Computing infrastructure in order to facilitate interoperability between vendor solutions led to the

creation of Utility Computing Working Group under the auspices of DMTF (Distributed Management Task Force) [5] and with the participation of major IT players including Cisco Systems, EDS, EMC, HP, IBM, Oracle, Sun Microsystems and VERITAS. The main objective of the DMTF Utility Computing Working Group is to develop a set of interoperability standards in collaboration with other organizations including OASIS (Organization for the Advancement of Structured Information Standards) and GGF (Global Grid Forum) that will allow the assembly of comprehensive services from components supplied by different vendor platforms.

3.3 Role of SOA and Web Services

Another key trend favouring application servicing over the traditional software-licensing model is the move towards service-oriented architecture (SOA) for enterprise computing. The nature of enterprise applications have changed dramatically over the last five years; most enterprise applications today have requirements to interoperate across enterprise boundaries (i.e. requirements for e-business). Service-oriented computing based on Web services standards and technologies is widely regarded as having the potential to address the requirements for e-business interoperability and are likely to become the dominant enterprise computing architecture in the future. There is a close relationship between application servicing and service-oriented computing. Web services are regarded as the enabling technology for the integration of ASP applications, and for delivery of low-granularity application services [8]. The wide adoptions of Web services standards across the various computing environments (i.e. .Net, enterprise Java) makes Web services an ideal solution for application integration, and for exposing selected business functions of complex enterprise applications.

In summary, business and technological factors discussed above have created a situation where delivery of enterprise application in the form of services becomes both technically possible, and economically desirable. This is likely to have major impact on enterprise computing over the next five years, finally tipping the balance from licensed software towards software delivered as a service.

4. COMPARISON OF THE SOFTWARE-AS-A-SERVICE MODEL WITH TRADITIONAL MODEL OF APPLICATION DELIVERY

ASP or the software-as-a-service model has crystallized into specific service delivery model in the last two years. In Table 1 we analyse the application service model from three perspectives: design and technology, business, and IT management viewpoints [10], [14], [15], [20], [21]. We compare the ASP model

with the traditional approach in which software vendor sells the software license and the customer runs the software on its own technology infrastructure. This analysis emphasizes the features of the ASP model that

can help organizations to solve their IS/IT problems in new and more efficient ways than the traditional approach.

Table 1: Comparison of the software-as-license vs. software-as-service models for enterprise applications

Differentiator	ASP (SW as Service)	Traditional Approach (SW as License)
Main characteristics	Application service provider controls all necessary IT resources and delivers application functionality to a large number of customers as a service via Internet. Many users from different organization at a time use the same application.	Software vendor develops the application; the application is implemented on customer's HW and customer is responsible for the operations.
Design and technology Issues		
Design premise	Designed from the outset for delivery as Internet-based service for a large number of customers. It includes specific HW and SW architectures, and business model.	Designed for implementation by specialist and for customer to operate and maintain.
Technological architecture	Multi-tenant service oriented architecture designed to run hundreds or thousands of users from different user organizations on a scalable technological infrastructure.	Architecture suitable for deployment by individual company on a dedicated IT infrastructure.
Client interface	Browser is the main and often the only interface for all applications. It eliminates the need to develop, install, and support multiple client interfaces.	Many SW vendors have added browser interfaces, but most support multiple clients – it increases development, installation and support costs.
Service management	Applications with embedded service management, monitoring metering and security capabilities.	Typically must add service management, monitoring and metering features subsequent to product development.
Upgrades	Frequent (every 3-6 month) upgrades possible. All customers are upgraded simultaneously resulting significant cost reductions.	Infrequent, major updates (every 12-24 months). Individual customers may be running different versions of software. Both, provider and customer, have to implement version management process.
Business Issues		
Readiness of the service	Short implementation cycle. Typically no requirement for new HW and SW	Long implementation cycle due to complex implementation of HW, SW, and knowledge transfer to customer sites.
Availability of the service	The service is available from any location (globally).	Could be limited to single organization via intranet or client/server interface.
Scalability of the service	The volume of the services delivered (i.e. number of users supported, number of transactions) can be scaled (up or down).	Configuration needs to support peak requirements, and cannot be scaled down.
Reliability of the service	Typically very high. Provider can more efficiently invest in network and systems redundancy. If he does not assure high reliability he loses customers' business.	It is very expensive to provision true fault-tolerance for in-house applications. Most companies remain at risk and typically experience periodic downtime.
Flexibility to business changes	Good if alternative service providers are available.	Good if the business change requires only minor application changes. Inflexible if the business change requires major application changes or new application development.
Customisation	Typically limited.	Extensive customisation possible (at both configuration and source-code levels), but expensive.
Functionality	Often limited functionality, application typically designed for narrow vertical market.	Extensive functionality, customers often use only small part of the available functions.
Evaluation of an application by user	The application can be evaluated before the purchase.	Application is evaluated after purchase, installation and customisation.
Internal sources utilization (people, technology, etc.)	Only few internal sources used for IT processes support. Most of the company sources can be used for core business processes.	Many internal sources used for IT processes support.
Costs of IT	Predictable, no investments required -operating costs only. The costs are highly correlated with the volume of services.	Both investments and operating costs. High overhead costs given by depreciation and amortizing of investments. The costs may not correlate with the volume of service delivered.
IT management Issues		

differentiator	ASP (SW as Service)	Traditional Approach (SW as License)
Subject matter of the contract	IT Service (SLA). The main parts of SLA: <i>content of the service</i> (functionality, data, training, support,...), <i>volume</i> (number of users, number of transactions, volume of data,...), <i>quality</i> (availability, response time, security,...) and <i>price</i> .	Usually, the contract divided into several subcontracts for hardware, SW licence, service (implementation, integration, training, upgrade,...).
SLA	The usage of SLA is a standard requirement	SLA in most cases not used.
Responsibility for IT infrastructure	Provider.	Customer (but some of the activities often realized by third parties).
IT sources utilization	IT sources of the provider (HW, SW, IT specialists) are used across all customers; provider has advantages of economies of scale. Customer IT sources are minimized.	IT sources are used only for one organization.
IT knowledge required at customer site	How to use IT for competitiveness enhancement, available services at IT market, SLA structure, and management of service delivery.	The same as in ASP plus: wide spectrum of IT knowledge. The required IT knowledge is dependent on number of platforms and types of application used. Extensive technical training needed
Size of IT personnel at customer site	Very small.	Large – different types of specialists needed.
Problem and change management procedures	Short feedback cycle - procedures enable almost immediate feedback. Support staff or programmers can directly identify and fix problems. Fixing a problem for one customer fixes it for everyone, which reduces support costs.	Problem solving is often indirect via intermediaries (VARs, SIs, etc). Patches and upgrades are implemented at individual customer sites. Costly and unreliable, as customers often delay installation of patches and upgrades.
Main risks	Loss of expertise that could be useful in the future. Stability of the provider (Exit strategy). Unsatisfactory customisation. Unresolved systems integration issues – who should be responsible for integration? Enhancements not under control of the customer. Security, satisfactory response time.	Stability of the provider – but not to the same extent as for ASP. Technology backwardness. High TCO (Total Cost of Ownership). Low flexibility and scalability.

The above table includes a number of compelling arguments that will make the software-as-service model a preferred solution for enterprise applications delivery in the future. Many of the benefits listed above are results of recent technology advances and increasing expertise of services providers.

Given the current ASP limitations suitability of applications for ASP delivery needs to be evaluated, as certain types of applications may not be suitable candidates for application servicing. More specifically, these applications include:

- [1] Mission critical core business applications – typically not available from external providers (the critical nature of these applications dictates in-house implementation and control),
- [2] Highly customized and specialized applications (providers can not gain economies of scale as the number of customers using such applications is very small),
- [3] Applications with extensive integration requirements (such applications have close dependencies on other enterprise applications and cannot be effectively managed externally).

On the other hand, standardized applications such as ERP or CRM are good candidates for application servicing.

5. THE IMPACT OF ASP ON THE IT LANDSCAPE

There are important consequences of the shift from the software-as-license model to the software-as-service model for the delivery of enterprise applications. The emergence of Utility Computing as the new paradigm for enterprise applications will have major impact on the IT landscape, creating new opportunities and challenges for both the providers and customer organizations. The reduction in the size of the traditional software license market, reduced demand for on-site implementation and the corresponding increase in demand for application services will lead to further rationalization of the IT vendor market [4].

The shift towards the service model for delivery of enterprise applications will take several years to be fully realized, but the impact on IT industry as a whole, and software vendors and third-party consulting organization in particular will be dramatic. It will result

in restructuring and re-alignment of the major industry players that will favor large software vendors who are already successfully delivering application services to end-user organizations. Consulting organizations whose main activity is implementing ERP systems and similar applications for client organizations will need to re-focus their business activities, as there not likely to be many large-scale implementation projects of this type in the future. The ASP model will be associated with new pricing models that are likely to dramatically reduce the cost of ownership of enterprise applications.

Reduction of demand for in-house IT specialists will lead to the restructuring of the IT labour market, and will demand important changes from user organizations that will need to implement suitable management structure and IT architecture that enable effective participation in the world of service-oriented computing.

6. CONCLUSIONS

In conclusion, the ASP or software-as-service model provides a viable alternative to traditional software licensing model for many application types today, and it is likely that the ASP model in combination with Utility Computing will become the dominant method for delivery of enterprise applications in not too distant future. *The main reason for it is that ASP model enables cheaper and more flexible IT service delivery than the traditional one.* This view is supported in the literature, for example Gartner ranked "software as service" as one of the current megatrends and predicts that up to 40 percent of all applications will be delivered over the Internet within the next 2 to 3 years.

REFERENCES

- [1] Aberbethy, M.: Outsource Tack Loses its Edge, *Newsweek Bulletin*, February 11, 2004, <http://bulletin.ninemsn.com.au/bulletin/EdDesk.nsf/All/FA57D33080F2C482CA256E2900778612>
- [2] Boulton, C.: IBM, Veritas Lead New Utility Computing Standard, <http://www.internetnews.com/ent-news/article.php/3311791>
- [3] Carr, N.G.: IT Doesn't Matter, *Harvard Business Review*, Vol. 81, No. 5, May 2003
- [4] Cohen, P.: Twelve Technical and Business Trends Shaping the Year Ahead, <http://www.babsoninsight.com/contentmgr/showdetails.php/id/687>
- [5] DMTF Announces New Working Group for Utility Computing, http://www.dmtf.org/newsroom/releases/2004_02_17
- [6] Dubie, D.: Vendors make the utility computing grade, *Network World Fusion*, 22/03/2004, <http://www.nwfusion.com/news/2004/0322dellsummit.html>
- [7] Eriksen, L.: Will the Real Utility Computing Model Please Stand Up, <http://www.utilitycomputing.com/news/342.asp>
- [8] Ferguson, R.B.: SAP Sets Its Sights on SOAP , 10/5/2004, http://www.eweek.com/print_article/0,1761,a=126512,00.asp
- [9] Feuerlicht G. and Phipps T, Architecting Enterprise Applications for the Internet, *Journal of Applied Systems Studies*, Special Issue on "Applied Cooperative Systems", July, 2000, ISSN 1466-7738.
- [10] Feuerlicht, G., Voříšek, J.: Key Success Factors for Delivering Application Services, *Proceedings of "Systems Integration 2003" conference*, VŠE, Praha, 2003, 274-282, ISBN 80-245-0522-3
- [11] Haber, L.: ASPs Still Alive and Kicking, January 30, 2004, <http://www.aspnews.com/trends/article.php/3306221>
- [12] IDC's Findings Reveal ASP Implementations Yielded an Average Return of Investment (ROI) of 404%, FRAMINGHAM, Mass., 8 February 2002, <http://www.idcindia.com/pressrel/Show.pressrel.asp?path=prt20020038.htm>
- [13] Karpecki, L.: Analýza koncového trhu - Aktuální stav a trendy 1999-2001, *INSIDE*, č. 2, 2002, pp.26-45
- [14] Levy, A.: What Do Enterprises Want From ASPs?, February 2004, http://www.aspnews.com/analysis/analyst_cols/article.php/2217411
- [15] McCabe, L.: A Winning Combination: Software-as-Services Plus Business Consulting and Process Services, *Summit Strategies Market Strategy Report*, <http://www.summitstrat.com/store/3ss07detail>
- [16] Nevens, M.: The real source of the productivity boom, *Harvard Business Review*, March, 2002, pp. 23-24
- [17] Never-ending Story Why ERP Projects Cause Panic Attacks, Survey of 117 companies by The Conference Board Inc. in NY, http://www.computerworld.com/cwi/story/0,1199,NA_V47_STO64064,00.html
- [18] Oracle Financial Reports, http://www.oracle.com/corporate/investor_relations/analysts/
- [19] SAP Financial Reports, <http://www.sap.com/company/investor/reports/>
- [20] Voříšek J., Pavelka J., Vít M.: Aplikční služby IS/IT formou ASP - Proč a jak pronajímat informatické služby, Graga Publishing, Praha, 2003, ISBN 80-247-0620-2, (Application Service Providing – Why and How to Hire Application Services)
- [21] Wainwright, P.: Secret Weapons of ASPs, February 2004, http://www.aspnews.com/analysis/analyst_cols/article.php/3090441

Comparison of E-Commerce Practices between Large Enterprises and Small to Medium Enterprises

Sehyung Cho¹, Yongkyun Chung², Seungchul Kim³

¹Dept of Business Administration Konyang University, Korea, shcho@kytis.konyang.ac.kr

²Div. of Econ. and Int. Trade, Kangwon National Univ, Korea, ykchung@kangwon.ac.kr

³College of Business Admin., Hanyang University, Korea, sckim888@hanyang.ac.kr

ABSTRACT

E-commerce becomes an important element of the business strategy for both small and medium enterprises (SMEs) and large enterprises as online business and cyber market space continue to expand its size and volume. E-commerce has a particularly important implication for SMEs because it can help SMEs overcome many handicaps of the SMEs, such as the economies of scale, and compete against the large enterprises on the level ground. The two groups of business firms, large enterprises and SMEs, are often different in not only their sizes but also their business practices. This may also extend to how they use e-commerce in their business. This paper is to study whether there are differences in using e-commerce between large enterprises and SMEs, and what are the differences and why the differences are made. We compare the two groups of business firms in two aspects, format of implementation and their performance. Data used for the study is collected from a group of Korean firms.

Keywords: Electronic commerce, Business strategy, Small firms

1. INTRODUCTION

Electronic commerce (EC) has caused many changes in the economic structure, and now begins to develop a new paradigm for business process (Kalakota and Whinston, 1996). Business firms are taking part in online business activities because they cannot survive without participating in e-commerce. Thus, e-commerce is now an essential part of business strategy for survival and growth (Kettinger and Hackbarth, 1997; Levy et al., 1999; Poon, 2000) for both large enterprises and small to medium enterprises (SMEs).

The research on e-commerce so far has some limitations. First, it lacks empirical study conducted with field data collected from business firms. Most of the past studies also focused on the issues of adoption and diffusion of e-commerce (Cragg and King, 1993; Fink, 1998; Mehrkens et al., 2001). E-commerce nowadays, however, is considered a must rather than a matter of choice. Thus, we need to discuss the issue of how to use e-commerce more effectively. Second, past studies tend to overlook the effect of the size of the firm on the implementation and the performance of e-commerce. There are clear differences in the way of conducting business such as marketing and operations between large enterprises and SMEs (Thong et al., 1996), and these differences would be reflected in the way of utilizing e-commerce and need to be studied.

The differences between large enterprises and SMEs can be summarized as follows: First is the issue of the firm size. Usually, large enterprises take advantage of the economies of scale by producing a few items in large quantities while SMEs tend to rely on the differentiation ability by producing a variety of items in small quantities. E-commerce is known to neutralize the effect of the firm size on the competition, and thought to help SMEs overcome the disadvantage of the size in the market. Especially the market presence, that is, the power of reaching a large number of customers, may be accomplished very easily even by a very small firm by using e-commerce. One of the examples may be a home-shopping company which advertises and sells its products on cable TVs. We would like to find out whether the firm size is really overcome by e-commerce and has no impact on the business results in the world of e-commerce and internet age.

Second, SMEs are small in size with a relatively small sunk cost, and thought to be able to adapt to the change of the market situation quickly compared to large enterprises (Poon and Swatman, 1998). This agility gives SMEs an advantage to be able to innovate fast in the fast changing business environment of today's world. Therefore, e-commerce is again considered to help SMEs more than large enterprises.

Third, large enterprises have advantages in human resource and capital, and therefore can implement

³ Corresponding author

e-commerce in a more comprehensive manner. Also, the level of specialty is considered higher than SMEs that tend to hire generalists and use e-commerce for more general purposes.

As discussed above, large enterprises and SMEs are known to have differences in many aspects, but there has not been enough empirical study with field data to find out whether these differences extend to the e-commerce practices. This paper is to study whether there are differences in using e-commerce between large enterprises and SMEs, and what are the differences and why the differences are made. We compare the two groups of business firms by using the data collected from a group of Korean firms.

2. RESEARCH METHODOLOGY

2.1 Research Model and Hypotheses

The research model in Figure 1 shows the overall conceptual relationships of the variables affecting the implementation of e-commerce and its performance. While many of the previous studies only focused on the adoption issues (Cragg and King, 1993; Fink, 1998), our model is more comprehensive in that all the factors that affect adoption, implementation, and performance of e-commerce are included in the research model. Table 1 contains the explanations of the variables

Figure 1. Research model for E-Commerce.

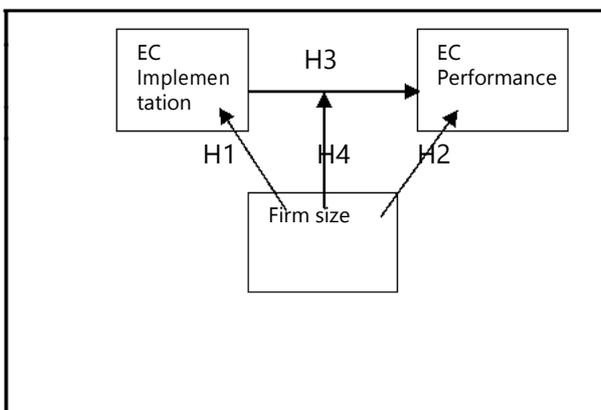


Table 1. Explanation of the variables.

Variables	Contents
Environmental variables	Firm size
Implementation variables	EC characteristics EC format EC strategy EC type
Performance variables	Utilization Satisfaction Usefulness

H1: Hypotheses on firm size and E-commerce

implementation:

Firm size is an indicator of a firm's resource and capability. Management decision by a firm is often constrained by its resource and capability. Thus, firm size will affect a firm's decision on EC implementation in such areas as the characteristics, format, strategy and type of E-commerce. Four hypotheses are tested and the details are listed in Table 2.

H2: Hypotheses on firm size and E-commerce performance:

The E-commerce will produce different results for the firms depending on the firm's size when other conditions are equal because of the economies of scale. Regardless of firm's size, E-commerce requires a certain amount of investment, and it is thought that the return on investment will be greater for the large firms than for the small firms. The details of the hypotheses are listed in Table 2.

H3: Hypotheses on E-commerce implementation and E-commerce performance:

The impact of the EC implementation on the firm's performance is measured. A firm's business performance will reflect the results of E-commerce which the firm implements. While it is generally accepted that E-commerce affects the performance and provides many opportunities for business firms, we need to know the direction and magnitude of impact by E-commerce on the performance. Implementation decision includes determining the characteristics, format, strategy and type of E-commerce for the firm, and the firm's performance is measured by utilization, satisfaction and usefulness of the E-commerce assessed by the firm's management. The details of the hypotheses are listed in Table 2.

H4: Hypotheses on the moderating effect of firm size:

A firm's performance is generally improved by the implementation of E-commerce. The magnitude of improvement, however, would be different by the firm size due to the difference in the cost structure, staff capability, knowledge level, and the area of usage of E-commerce. Thus, it is thought that the firm size will have a moderating effect on the relations between E-commerce implementation and performance. The details of the hypotheses are listed in Table 2.

Table 2. Hypotheses.

Variables	Hypothesis	Content
Firm size and EC implementation	H1-1	Firm size affects the characteristics of EC that a firm implements.
	H1-2	Firm size affects the format of EC that a firm implements.
	H1-3	Firm size affects the strategy of EC that a firm implements.

	H1-4	Firm size affects the type of EC that a firm implements.
Firm size and EC Performance	H2-1	Firm size affects EC utilization.
	H2-2	Firm size affects EC satisfaction.
	H2-3	Firm size affects EC usefulness.
EC implementation and EC performance	H3-1	EC characteristics has an impact on EC performance.
	H3-2	EC format has an impact on EC performance.
	H3-3	EC strategy has an impact on EC performance.
	H3-4	EC type has an impact on EC performance.
Moderating effect of Firm size	H4-1	The impact of EC characteristics on EC performance is affected by firm size.
	H4-2	The impact of EC format on EC performance is affected by firm size.
	H4-3	The impact of EC strategy on EC performance is affected by firm size.
	H4-4	The impact of EC type on EC performance is affected by firm size.

2.2 Data Collection and Analysis

We have circulated survey questionnaires to collect data from business firms in Korea. A total of 124 firms returned the questionnaires, and 99 of these were used for analysis after excluding those with missing data. The profile of the respondents are as follows (Table 3). Statistical tests such as cross tabulation, factor analysis, analysis of variance, and t-test are used to analyze the data and to find out the answers for the hypotheses presented in the above.

Table 3. Profile of respondent firms.

Type of industry	Number of firms	Proportion
Manufacturing	40	40.4%
Sales/Distribution	20	20.2%
IT/Technology	23	23.2%
Service	16	16.2%
Total	99	100%

3. ANALYSES AND RESULTS

The reliabilities are tested for the measured variables, and the results in Table 4 (cronbach $\alpha > 0.9$) show that all of them are well above the acceptance level. Also, factor analysis is conducted to confirm the validity of the performance variables. The 18 items in the questionnaire used to measure E-commerce performance were factor analyzed using principal component analysis with varimax rotation. We used

eigenvalue of one and the factor loading of 0.5 as the selection criteria. This yielded three factors with 3 items for each factor, as shown in Table 5. The three factors are named as usefulness, satisfaction, and utilization.

Tables 6 through 9 show the results of the tests to find out the relationship among the variables and to determine the acceptance of the hypotheses presented in the above.

Table 4. Reliability test.

Variables	Number of items	Cronbach α
IS maturity	9	0.9096
EC utilization	6	0.9324
EC satisfaction	6	0.9236
EC usefulness	6	0.9550

Table 5. Results of the factor analysis for the performance variables.

Question items	Factors		
	Factor 1 utilization	Factor 2 satisfaction	Factor 3 usefulness
Hours used	<u>.874</u>	.208	.244
How often	<u>.805</u>	.220	.355
How much	<u>.738</u>	.263	.383
Satisfaction	.119	<u>.794</u>	.321
Work support	.212	<u>.848</u>	.285
Ease of use	.334	<u>.845</u>	.146
Advantage	.362	.334	<u>.807</u>
Business results	.333	.246	<u>.861</u>
Improving	.345	.302	<u>.811</u>

3.1 Firm size and E-commerce Implementation

The hypotheses about the effect of the firm size on the firm's decision on E-commerce implementation are tested by cross-tabulation analysis. The results are presented in Table 6. Four dimensions of the firm's implementation decisions are analyzed: the characteristics, format, strategy, and type of E-commerce.

Firm size does not affect any of the E-commerce implementation decisions. For characteristics, all firms tend to use E-commerce for value system, that is, to enhance transaction values. Also, firm size does not affect the format of E-commerce significantly. Firms, however, appear to have different formats for their E-commerce, that is, large firms have more structured E-commerce format, and small firms have less structured format. In terms of strategy, firms do not adopt different strategy by its size. Lastly, firm size does not affect the firm's decision on E-commerce type selection. Both large and small firms adopt Business-to-Consumer (B2C) more often.

Table 6. Impact of firm size on EC implementation.
(Insert Table 6 here, available upon request)

3.2 Firm size and E-commerce Performance

We tested the relations between firm size and the firm's E-commerce performance measured in utilization, satisfaction and usefulness. The t-test results in Table 7 show that firm size does not significantly affect any of the performance measures. Thus, firms or its staff and employees do not feel the difference in E-commerce performance based on the firm size. In other words, the staff of small firms at least feel the same degree of satisfaction and usefulness about E-commerce, and utilize it as much as those of large firms.

Table 7. Impact of firm size on EC performance (t-test).

Performance variable	Firm size	Means	results
EC Utilization	Small	13.4769	t = 0.416 p = 0.678
	Large	13.0606	
EC Satisfaction	Small	12.7273	t = -1.542 p = 0.126
	Large	13.6061	
EC Usefulness	Small	12.2769	t = -1.014 p = 0.313
	Large	13.0909	

3.3 E-commerce implementation and performance

We tested the relations between the four E-commerce implementation variables discussed above (characteristics, format, strategy, and type) and the EC performance experienced by the firm measured in utilization, satisfaction and usefulness. The results of t-tests and analysis of variance (ANOVA) are presented in Table 8.

Some of the implementation variables have significant impact on the firm's E-commerce performance. Two implementation variables of E-commerce characteristics and format, significantly affect all three performance measures at the significance level of $\alpha=0.05$. E-commerce type has significant impact only on the utilization out of three performance measures. E-commerce strategy does not affect the performance at all. That is, firms do not show any difference in performance based on the strategy they adopted for E-commerce.

Table 8. Impact of EC implementation on EC performance

(Insert Table 8 here, available upon request)

3.4 Moderating effect of firm size

We conducted ANOVA tests to see whether there is a moderating effect of firm size on the relations between E-commerce implementation and performance. The

results are presented in Table 9. To sum up, the moderating effect of firm size was found to be significant for only a few cases.

For small firms, firm size has influenced only the satisfaction measure of the performance when E-commerce format and strategy are concerned. For large firms, firm size has more influence on the relations between the E-commerce implementation and performance. Firms show large difference in E-commerce performance depending on what they decide on the implementation issues of characteristics, format, strategy, and type. These factors in fact define the shape of the E-commerce system of a firm, and the results imply that the impact of the E-commerce system design is relatively more important for large firms than for small firms.

Table 9. Moderating effect of firm size on the relations between EC implementation and EC performance (ANOVA).

(Insert Table 9 here, available upon request)

4. CONCLUSION

We have found that the impact of firm size is not significant for most cases of E-commerce implementation decisions and performance. Firm size has moderating effect on the certain cases of E-commerce implementation and performance. E-commerce are said to help small-and-medium enterprises (SMEs) compete against large firms on a more level ground by moving the competition from the physical market place to cyber market space, and thus reducing the marketing and distribution cost. From the results, it seems that small firms, at least, use E-commerce as extensively as large firms, and enjoy the same degree of satisfaction and usefulness. Thus, we can say that E-commerce helps small firms enhance their operations up to the level of large firms, if not to overcome the firm size disadvantage completely.

This study has been conducted with a relatively small sample size, especially, of large firms. We need to continue to collect data from more firms, and then the study will produce more reliable and meaningful results.

5. REFERENCES

[1] Blili, S. and Raymond, L. (1993) Information Technology: Threats and Opportunities for Small and Medium-sized Enterprises, *International Journal of Information Management*, vol. 13, no. 6, pp. 439-448.
 [2] Brynjolfsson, E. (1993) The Productivity Paradox of Information Technology: Review and Assessment, *Communications of the ACM*, vol.35, no.12, pp.67-77.
 [3] Cragg, P. B. and King, M (1993) Small-Firm Computing: Motivators and Inhibitors, *MIS Quarterly*, vol. 17, no. 1, pp. 47-60.

- [4] DeLone, W. H. (1981) Firm Size and the Characteristics of Computer Use, *MIS Quarterly*, vol. 5, no. 4, pp. 65-77.
- [5] DeLone, W. H. and McLean, E. R. (1992) Information System Success: The Quest for the Dependent Variable, *Information System Research*, vol. 3, no. 1, pp. 60-95.
- [6] Doukidis, G. I., Smithson, S., and Lybereas, T. (1992) Approaches to Computerization in Small Businesses in Greece, *Proceedings of the Thirteenth ICIS*, pp. 139-148.
- [7] Fink, D. (1998) Guidelines for the Successful Adoption of Information Technology in Small and Medium Enterprises, *International Journal of Information Management*, vol. 18, no. 4, pp. 243-253.
- [8] Goode, S. and Stevens, K. (2000) An Analysis of the Business Characteristics of Adopters and Non-Adopters of World Wide Web Technology, *Information Technology and Management*, no. 1, pp. 129-154.
- [9] Grover, V. and Ramanlal, P. (1999) Six Myths of Information and Markets: Information Technology Networks, Electronic Commerce, and the Battle for Consumer Surplus, *MIS Quarterly*, vol. 23, no. 4, pp. 465-495.
- [10] Igbaria, M., Zinatelli, N., and Cavaye, A. L. M. (1998) Analysis of Information Technology Success in Small Firms in New Zealand, *International Journal of Information Management*, vol. 18, no. 2, pp. 103-119.
- [11] Kettinger, W. J. and Hackbarth, G. (1997) Selling in the Era of the Net: Integration of Electronic Commerce in Small Firms, *Proceedings of the Eighteenth ICIS*, pp. 249-262.
- [12] Kowtha, N. R. and Choon, T. W. I. (2001) Determinants of Website Development: A Study of Electronic Commerce in Singapore, *Information & Management*, vol. 39, no. 3, pp. 227-242.
- [13] Lai, V.S. (1994) A Survey of Rural Small Business Computer Use: Success Factors and Decision Support, *Information & Management*, vol. 26, no. 6, pp. 297-304.
- [14] Lederer, A. L., Mirchandani, D. A., and Sims, K. (2001) The Search for Strategic Advantage from the World Wide Web, *International Journal of Electronic Commerce*, vol. 5, no. 4, pp. 117-133.
- [15] Lee, H. G. and Clark, T. H. (1997) Market Process Reengineering through Electronic Market Systems: Opportunities and Challenges, *Journal of Management Information Systems*, vol. 13, no. 3, pp. 113-136.
- [16] Leidner, D. E. (1999) Virtual Partnerships in Support of Electronic Commerce: the Case of TCIS, *Journal of Strategic Information Systems*, no. 8, pp. 105-117.
- [17] Levy, M., Powell, P., and Galliers, R. (1999) Assessing Information Systems Strategy Development Frameworks in SMEs, *Information & Management*, vol. 36, no. 5, pp. 247-261.
- [18] Mehrtens, J., Cragg, P. B., and Mills, A. M. (2001) A Model of Internet Adoption by SMEs, *Information & Management*, vol. 39, no. 3, pp. 165-176.
- [19] Nath, R., Akmanligil, M., Hjelm, K., Sakaguchi, T., and Schultz, M. (1998) Electronic Commerce and the Internet: Issues, Problems, and Perspectives, *International Journal of Information Management*, vol. 18, no. 2, pp. 91-101.
- [20] Nunnally, J. C. (1978) *Psychometric Theory*, 2nd ed., McGraw Hill, New York, NY.
- [21] Palvia, P.C. (1997) A Model and Instrument for Measuring Small Business User Satisfaction with Information Technology, *Information & Management*, vol.31, no.3, pp.151-164.
- [22] Poon, S. (2000) Business Environment and Internet Commerce Benefit - A Small Business Perspective, *European Journal of Information Systems*, no. 9, pp. 72-81.
- [23] Subramani, M. and Walden, E. (2001) The Impact of E-Commerce Announcements on the Market Value of Firms, *Information Systems Research*, vol. 12, no. 2, pp. 135-154.
- [24] Thong, J. Y. L., Yap, C. S., and Raman, K. S. (1996) Top Management Support, External Expertise and Information Systems Implementation in Small Businesses, *Information Systems Research*, vol. 7, no. 2, pp. 248-267.

Demonstrative Analysis of Hi-tech Enterprises Technology Innovation Capability Based on ISM

Yu Chang¹, Naiding Yang¹, Xiandong Liu²

¹ Management School, Northwestern Polytechnical University, Xi'an 710072, China

² Aerospace Times Electronics Corporation
{changyu, naidingy}@nwpu.edu.cn

ABSTRACT

This paper presents a scientific exploration of the analysis and assessment of high-tech enterprises' technologic innovation capabilities, by applying the method of system engineering and ISM. Combining with enterprises' practice of XI'AN high-tech development zone, it focuses on the application of these methods and the results analysis. Some proposals and solutions for management are also raised in the paper.

Key words: High-tech enterprise, Technology innovation capability, Interpretative Structural Modeling

1. INTRODUCTION

The analysis and evaluation of technology innovation capability for hi-tech enterprises is a important task for finding the problem, improving and boosting the level of technology innovation capability. Due to the complex of hi-tech enterprises' technology innovation capability, it is difficult to quantify the analysis and evaluation of technology innovation capability. So managers are facing a problem for finding a analysis method which can reflects hi-tech enterprises high technology innovation capability. In order to solve the problem, the paper adopts ISM to analyze and evaluate hi-tech enterprises' technology innovation capability.

2. SUMMARIZED ACCOUNT OF ISM

Interpretative Structural Modeling (ISM) is one of structural modeling technique developed by John Warfield in order to analyze some problems of complex social economy system. The characteristic is dividing complex system into some system elements, depending on the people's practice experience, knowledge and computer forming a hierarchical structural model. The working procedure of ISM is as follows:

(1) Found an ISM group. There are about 10 people for every group and everyone care about the problem needed to solve. Besides, the people must have the different ideas, if there is a person who can make policy timely in the group, the discussion will be more efficient. There are often three kinds of people taking parting the group and they are usually : Method technique experts, Coordinators and Participators. (2) Determin the key problem. (3) choose the causative factors which comprise the system and affect the key problem. (4) List the

correlation of causative factors. (5) According to the structural model, establish ISM.

3. SPECIFIC IMPLEMENTATION AND DEMONSTRATIVE ANALYSIS

3.1 Founding ISM group

There are more than 10 members in the ISM group including the leaders of Xi'an hi-tech industry development zone productivity promotion center, foundation management center, economy development bureau and the professors of northwestern polytechnical university.

3.2 Corroborate the key problem and causative factors, list the correlativity of each causative factors

Hi-tech industry development zone, corroborating the problem is the reasons of hi-tech enterprises' weak technology innovation capability, analyze the 12 causative factors (table 1). Members of group clarify the relationship between the 12 causative factors after many times' discussion and fill the table 2 according to the influential relationship as follows.

- (1) If S_i has influence on S_j , fill 1; whereas fill 0 ($i, j=0, 1, \dots, 12$)
- (2) For the mutual influential factors, choose the more influential factor as the influence relationship, i.e. having the influence

3.3 Establish reachcapability matrix

Table1 causative reasons

Key problem: reasons of sick hi-tech capability for hi-tech enterprises		S ₀
Causative reasons		
1	pay more attention to present benefit, overlook long-term benefit and fostering of technology innovation	S ₁
2	the separateness of technology strategy and enterprise development strategy	S ₂
3	lack of technology investment and improper investment	S ₃
4	value product innovation, overlook engineering innovation	S ₄
5	lack of continuous strategy core in choice and management of project	S ₅
6	overlook the influence of culture	S ₆
7	lack of scientific management in technology innovation	S ₇
8	Weak capability of marketing	S ₈
9	separateness of research and production, low efficiency of R&D	S ₉
10	lack of cooperation between enterprise and university, enterprise and research department	S ₁₀
11	impediment of information communication, waste of R&D resource	S ₁₁
12	ineffective innovation policy	S ₁₂

Table2 experts indgement relationship table (blank means 0)

	S ₀	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈	S ₉	S ₁₀	S ₁₁	S ₁₂
S ₀	1												
S ₁	1	1		1	1			1		1	1	1	1
S ₂	1		1	1	1			1		1	1	1	1
S ₃	1			1									
S ₄	1				1								
S ₅	1			1	1	1		1	1	1		1	
S ₆	1			1	1		1	1		1		1	
S ₇	1			1	1			1		1		1	
S ₈	1								1	1			
S ₉	1									1			
S ₁₀	1										1		
S ₁₁	1											1	
S ₁₂	1									1	1		1

3.4 Give the reachcapability matrix a multi-level partition and establish structure model

(1) Explanation relevant concepts

reachcapability set: the set of elements that can be reached by element S_i is defined as the reachcapability set of element S_i, expressed by K(S_i).It's made up of the elements whose value is 1 in row S_i of reachcapability matrix.

Origin set: The set of elements that can reach element S_i is defined as the origin set of element S_i, expressed by A(S_i). It's made up of the elements whose value is 1 in column S_i of reachcapability matrix.

The most advanced element set: The most advanced element set of multi-level and increasing degree structure, refers to the set whose element can't reach other

elements except itself. Its reachcapability set only include the element of itself, but in origin set, except that including element S_i, it also includes the next-level elements that can reach it.

If $K(S_i) = R(S_i) \cap A(S_i)$, here $i=j$, then R(S_i) is the most advanced element set.

After finding out the most advanced element set, we can cross out the corresponding rows and columns in reachcapability matrix. Then, go on looking for the most advanced element in the leaving matrix. Go on as above, we can find out the most advanced element set included in each level.

(2) Looking for each advanced element set.(as table3 shown)

This degree only $R(S_0) \cap A(S_0) = R(S_0)$, so the most advanced element is 0 for this degree, and the first

degree element is $\{S_0\}$, cross out the corresponding rows and column, then get the second degree reachcapability set and origin set . Going on as the above, we can

similarly get the second elements set $\{S_3 S_4 S_9 S_{10} S_{11}\}$, then, the third elements set $\{S_7 S_8 S_{12}\}$, finally, the deepest elements set $\{S_6 S_2 S_5 S_1\}$.

Table3.the first reachcapability set and origin set

S_i	$R(S_i)$	$A(S_i)$	$R \cap A$
S_0	0	0,1,2,3,4,5,6,7,8,9,10,11,12	0
S_1	0,1,3,4,7,9,10,11,12	1	1
S_2	0,2,3,4,7,9,10,11,12	2	2
S_3	0,3	1,2,3,5,6,7	3
S_4	0,4	1,2,3,4,5,6,7	4
S_5	0,3,4,5,7,8,9,11	5	5
S_6	0,3,4,6,7,9,11	6	6
S_7	0,3,4,7,9,11	1,2,5,6,7	7
S_8	0,8,9	5,8	8
S_9	0,9	1,2,5,6,7,8,9,12	9
S_{10}	0,10	1,2,10,12	10
S_{11}	0,11	1,2,5,6,7,11	11
S_{12}	0,9,10,12	1,2,12	12

(3) Reachcapability matrix divided by levels

	S_0	S_3	S_4	S_9	S_{10}	S_{11}	S_7	S_8	S_{12}	S_1	S_2	S_5	S_6
S_0	1	0	0	0	0	0	0	0	0	0	0	0	0
S_3	1	1	0	0	0	0	0	0	0	0	0	0	0
S_4	1	0	1	0	0	0	0	0	0	0	0	0	0
S_9	1	0	0	1	0	0	0	0	0	0	0	0	0
S_{10}	1	0	0	0	1	0	0	0	0	0	0	0	0
S_{11}	1	0	0	0	0	1	0	0	0	0	0	0	0
S_7	1	1	1	1	0	1	1	0	0	0	0	0	0
S_8	1	0	0	1	0	0	0	1	0	0	0	0	0
S_{12}	1	0	0	1	1	0	0	0	1	0	0	0	0
S_1	1	1	1	1	1	1	1	0	1	1	0	0	0
S_2	1	1	1	1	1	1	1	0	1	0	1	0	0
S_5	1	1	1	1	0	1	1	1	0	0	0	1	0
S_6	1	1	1	1	0	1	1	0	0	0	0	0	1

±±

(4) Found of structure model(figure 1)

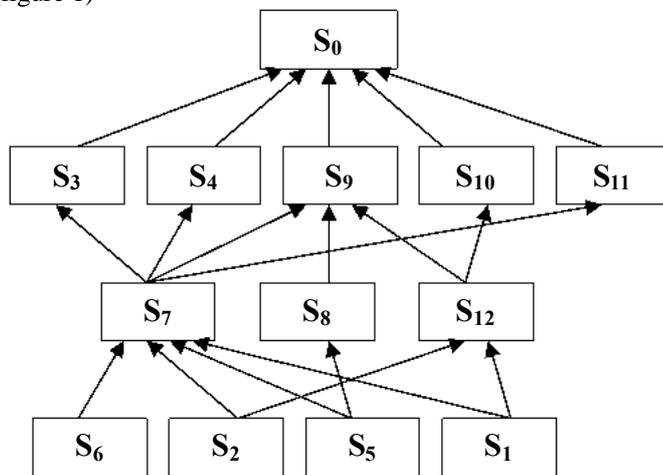


figure1 structure model

3.5 Establishing ISM according to structure model((figure 2)

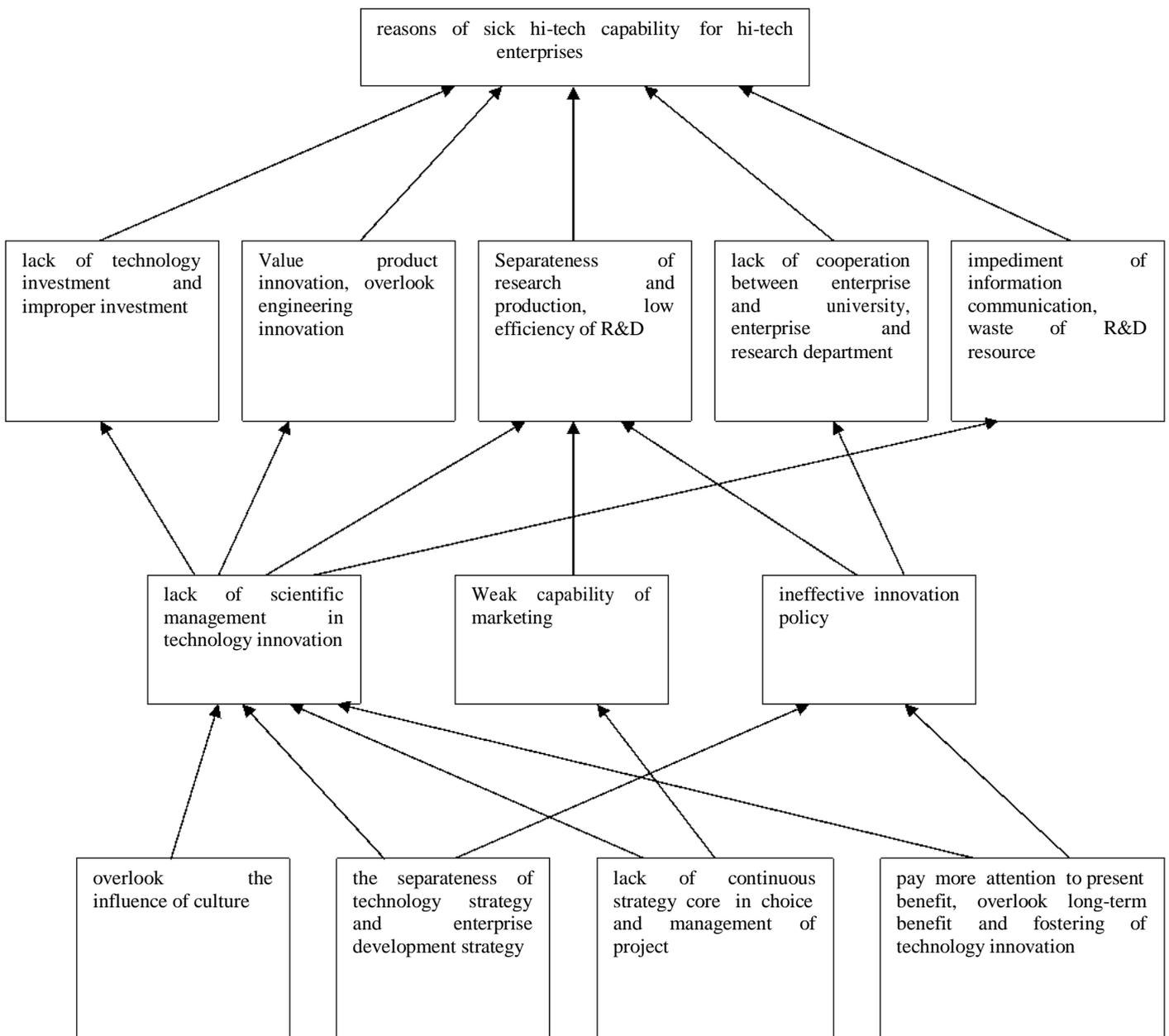


figure2 ISM of hi-tech enterprises technology innovation capability weakness

4. ANALYSIS OF TECHNOLOGY INNOVATION CAPABILITY WEAKNESS BASED ON ISM

According to figure 2 , we know hi-tech enterprises technology innovation capability have four levels, the four causative reasons of the lowest level are: (1) overlook the influence of culture. (2) the separateness of technology strategy and enterprise development strategy. (3)lack of continuous strategy core in choice and management of project. (4)pay more attention to present benefit, overlook long-term benefit and fostering of technology innovation.

Peter.Drucker thinks innovation as a important pillar of

modern enterprise culture. If innovation culture becomes the enterprise's basic characteristic, innovation values will get all staffs' agreement, innovation system and behavior standard will be set up, motivation and operation mechanism of enterprise's technology innovation will form and operate efficiently. But hi-tech enterprises often overlook the influence of culture and organization, are unable to create culture atmosphere and environment which is benefit for technology innovation and lack of effective value orientation and motivation mechanism. Because some enterprises are lack of capability of making strategy and culture adjustment for changeable market, leading them to dilemma in market competition finally. For hi-tech enterprises, we must

combine the technology innovation strategy with culture development strategy. If they are separated, will lead to the low R&D efficiency, make research and production separation, at the lowest enterprise technology innovation activity can't sufficiently promote the economic development. Modern enterprise's overall strategy is resource employment methods of enterprise hierarchy, planning for enterprise's global and long-term exist and development. Overall management strategy is the premise of technology innovation. Only determining the type of overall management strategy, we can decide the basic methods of technology innovation. Enterprises core competition and technology is the core of continuous strategy. Enterprises must pay attention to the fostering and promotion the capability which can provide the long-term competitive advantage. If lacking of continuous strategy core on project's choice and management, paying attention to the present benefits, overlooking long-term benefits separating the technology connection, will lead to get no project package benefits. Lacking of systemic and packaged strategic ideas will baffle the development of technology innovation capability.

5. CONCLUSION

The example analysis of hi-tech enterprises technology innovation in hi-tech zone is accord with status quo of present hi-tech enterprises technology innovation capability. In order to foster and improve the technology innovation capability, hi-tech enterprises need to make the technology innovation capability as a system, analyze from whole angle, treat every sector seriously, especially solve the adverse factors. The main is: (1) paying attention to organization and culture influence (2) technology strategy is accord with enterprise development strategy (3) having the continuous strategy core in project choice (4) thinking highly of long-term

benefits and fostering of technology innovation capability.

Hi-tech enterprises only grasp the most basic factors which influence technology innovation capability, upswing and improve continuously, will ameliorate and improve technology innovation capability, overcome the character of sick technology innovation capability, accordingly improve the enterprises' competition capability, always make the enterprises as victors.

REFERENCES

- [1] Wang Yingluo, "Systematic Engineer Theory Method and Appliance", High Education Press, 1992:49-72 (in chinese)
- [2] Larry E. Westphal, Yung W. Rhee and Garry Pursell, "Sources of Technological Capability in the Third World," Edited by M. Fransman and K. King, 1984:163-279
- [3] Dorothy Leonard-Barton, "Wellsprings of Knowledge: building and sustaining the sources of innovation," Harvard business school press, Boston, Massachusetts, 1995: 170-320
- [4] Fu Jiaji, "Technology Innovation," Tsinghua University Press, 2000:10-50 (in chinese)
- [5] Ma Xiaoyong, Guang Jiancheng, "Policies Research For Improving Hi-tech Enterprises Technology Innovation Capability," Journal of Hua Bei Dian Li University, 2000:(2)4-7
- [6] Sun Yiming, "Modern Enterprises Technology Innovation". Shanxi Economic Press, 1988:20-40
- [7] Chang Yu, Lu Shangfeng, Liu Xiandong, "On Establishing Technological Innovation Capability Appraisal Index of HIDZ," Science and Technology Management Research, 2004(1):103-106 (in chinese)

Determinants of E-Commerce Adoption in SMEs

Pham T.T. Hong¹, Nguyen B. Giang²

¹ International Graduate School of Social Sciences, Yokohama National University,
Yokohama, 240-8501, Japan
d021c902@ynu.ac.jp or ptth11@yahoo.com

² Institute of World Economics and Politics, Vietnam Academy of Social Sciences, Hanoi, Vietnam
giangngb@yahoo.com

ABSTRACT

A deep insight of determinants of e-commerce adoption process is in needed to help small- and medium-sized enterprises (SMEs) to fasten up their e-commerce rush and overcome a great number of existed obstacles. This paper analyzed and classified factors taking effects on SMEs'e-commerce adoption into four groups, specifically: (1) perceived benefits, (2) organizational readiness, (3) external environment, and (4) leadership attitudes and competences. Especially, the paper also gets an overview on the role of leadership in motivating the uptake of e-commerce within SMEs. Thorough analysis within the research contributes not only a general view about core determinants of e-commerce adoption, but also the engine toward the success of e-commerce distribution channel in SMEs.

Keywords: E-commerce adoption, Determinants, Distribution channel, E-channel, and SME

1. INTRODUCTION

It has been observed and verified in many academic studies that SMEs have been actively looking for suitable solution and methods of adopting and integrating e-commerce into their business process. Based on different sense, researchers explored various factors that impact on SMEs in e-commerce adoption process. Focused on internationalizing SMEs, Tiessen (2001) suggested that SMEs'adoption e-commerce is driven both by market changes and by institutional norms. Following different approach, based on the Rogers (1995) model of innovation decision process, Kendall et al (2001) revealed that the compatibility issue is significant because it deals with their perception of the importance of e-commerce on their business now and in the future. Whereas, Levy and Powell (2001) demonstrated that the introduction of ICT into SMEs is fragmented and based on the operational support and transactions processing. Based on the discussion, they argued that a segmentation model, contingent on SMEs characteristics is applicable and may bring benefit for SMEs. Other existing researches indicated various factors appearing as enablers or limiters of e-commerce adoption in SMEs. Overall, they have been explored via interviews or surveys of SMEs implemented around the world (Poon and Swatman, 1999; Mehterns et al, 2001; Jeffcoate et al, 2002; Quayle, 2002; Lal, 2002; Olson and Boyer, 2003; Doherty et al, 2003; Drew, 2003; Pham, 2003). These previous researches showed a tendency for discussions about e-commerce to ignore the wider concerns involving distribution networks and hence the potentially crucial synergies.

Stemming from gaps in existing research, this study is aiming to answer the remaining question of "What are determinants of e-commerce adoption as novel distribution channel selection of SMEs?" During the

road, the study also portrays an in-depth insight in factors standing for the motivation leading a new e-channel design process distinguished from conventional ones, and consequently contribute a general view of an effective way to handle e-channel.

Section 2 of the paper is to discuss the whole picture of key factors affecting the e-commerce uptake process in a model of four components. Although some determinants composing these four components are widely recognized in the literature, the contribution of this paper is to synthesize the existing literature and adjust some factors including the role of government support, role of cultural factors, and the role of leadership attitudes into composing a conclusive model of determinants of e-commerce adoption. At the same time, the paper also suggests how to improve the insight of this important field in case of adopting e-commerce as a new digitalized distribution channel within SMEs.

The next part of the paper –section 3 –further discusses the important of considering these all components in a conclusive picture to accelerate SMEs to make a reasonable selection of electronic distribution channel.

2. DETERMINANTS OF E-CHANNEL ADOPTION AND E-CHANNEL DESIGN PROCESS

In general, enterprises may alter the most applicable model of e-commerce distribution channels based on three categories factors: (1) perceived benefits; (2) organizational readiness; and (3) influences external environment on the adoption decision process. However, for case of SMEs, the adoption of e-commerce model may depend on one more particular determinant –considered as the most important factor - the innovative attitude and knowledge about IT of the leadership. Figure 1 depicts determinants of SMEs'

e-commerce channel adoption motivation.

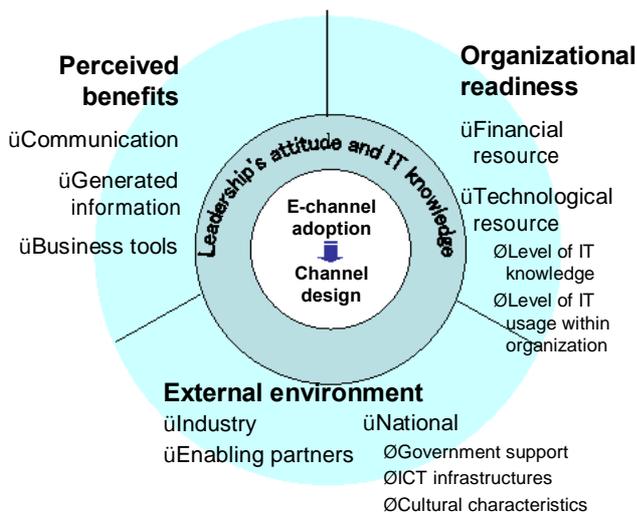


Figure 1

Factors impact distribution e-channel adoption

Source: Author's original work

2.1 Perceived benefits

The analytical results of Pham (2003) revealed that the more firms perceived the benefits they may gain from e-commerce implementation, the more firms are enthusiastic to engage e-commerce. In another words, as with many innovations, perceived benefits was often the characteristic that most influence its adoption and then affect their channel selection in fact.

From a theoretical point of view, and including the analysis of cases in which the e-commerce adoption has been successful, it is possible to identify a vast range of benefits that can be attained by SMEs companies via the e-commerce adoption. To get deep insight about the benefits that SMEs may gain from the so-called e-commerce, it is better to discover potential benefits into three groups, specifically: (1) communication, (2) generated information, and (3) business tools.

Enterprises experienced much of relative advantages from Internet as new methods of **communication** in comparison with traditional ones, such as telephones and faxes. With the support of information technologies, firms will easy to smooth up the cooperation between customer and supplier, and then, lead to the elimination of many causes of conflict and mistakes, and at the same time, lead to a closer working relationship. This allows supplier to concentrate on composing the product to meet the customers' needs and at the same time, encouraging these customers to focus on fewer suppliers as well. As the results, SMEs are enabled to maintain the central order and administration systems, and consequently, consolidate individual orders to regional loads.

It is easily identifiable and realizable that information

and communication technologies contribute much on **generating information**. The ability of generating information basically rewards a saving in costs regarding to document management. ICT allows firms to avoid re-inputting the data into the computer, reconciliation of documents, expenses of sending the documents by postal mail or fax, and cost related to filing and storing document including the space devoted to this. Moreover, ability of generating information network over all global allows SMEs to stockholding at central warehouse, and then, give a promise of comparatively low stockholding level.

The Internet was also viewed as a **business tool** and should be used as an integral part of every business day to support business objectives. The ability of direct delivery and installation to customer from regional warehouse speeds up the local delivery system. Besides, the ability to saving costs in the management of documents, ability to improve the efficiency of the internal organization of the firm and change its relationship with customers, ability to achieve greater quantity of generated information are closely related to the usage of the so-called e-commerce; and then, stand for major motivation toward diffusing e-commerce within enterprises over all the world.

2.2 Organizational readiness

Many organizational level variables such as centralization, size, complexity, slack, formalization, integration, functional differentiation, and specialization have been studied within the context of innovation diffusion (Kimberly and Evanisko, 1981; Damanpour, 1987; Meyer and Goes, 1988). These variables are important variables capturing the organizational context in which strategic decisions are made. However, in respect of e-commerce adoption, this study is essential to capture organizational readiness under two dimensions: the level of (1) financial and (2) technological resources of the firm.

Financial readiness refers to financial resources available for ICT to pay for installation costs, implementation of any subsequent enhancements, and ongoing expenses during usage. Usually, small firms with available financial resources will be better equipped to implement e-commerce. Consequently, firms that can afford more costly, e-commerce projects are more likely to enjoy higher benefits from the use of such systems. SMEs with enough equipped financial resource are more enthusiasms to adopt e-commerce as a distribution channel.

The second dimension of organizational readiness – **technological readiness** – is concerned with the level of sophistication of information technology (IT) usage and IT management in an organization. It is common that different companies have different processing needs and different information and communication technology

(ICT) infrastructures. It can be expected that the likelihood of adopting new information technologies or using IT in new business activities depend on the IT dependency or intensity of a company. For instance, the more a business depends on computerized information processes, the more likely the company is to be interested in a new type of IT usage.

The evidence from the survey on Japanese SMEs (Pham, 2004) supported the idea that the Internet was adopted by organizations that have the personnel who understand well about the technology. The more people within a company equipped relevant knowledge of information technology, the speedier firm accelerate to engage e-commerce.

2.3 External environment factors

(a) Industry

It is commonly accepted that distinguished characteristics of each industry facilitate its participants to adopt innovation in all manners: technological innovations, product innovations, managerial innovations and so forth. Regarding impacts of industrial characteristics on innovation adoption decision of the involving firms, inter-industry competition, product characteristics, and customers' traits are considering as major accelerators.

Competition from firms frequently involves innovation and subsequently promotional activity. When the innovations involve new communication technologies, their adoption enhances the overall communications infrastructure which, in turn, facilitates information innovations, particularly of new communication technologies. In this way, market competition has the unexpected effect on accelerating the diffusion of this new communication technology, particularly in regarding usage of Internet as an alternative distribution channel.

Beside the competitive status within the industry considered as an engine of e-commerce adoption intention of SMEs, **product features** are also considering as another major source to accelerate the adoption e-commerce as a new distribution channel within enterprises in all sizes. Product features are firstly considered as company has an intention of joining new economy because there are a lot of product is quite fit with e-commerce and performing e-commerce may accelerate the market share of this product in the market such as music, movie, software, and so forth. However, other product, in contrast, is in low level of relationship with e-commerce activities such as health care and beauty service.

Customers are always the major considerations of firms in designing appropriate distribution channels. In the information age, tougher than ever, customer acceptance and behavior become factors determining the

transformation of enterprises' distribution channels. The fact is that, for such a paradigm shift to be accepted by customers, firms have to develop resources and competencies that add value for customers. Rapid growth of customer acceptance has been witnessed recently, especially for websites assuring security and privacy, building with familiar interfaces, and providing rich information with reasonable download time.

Another particular characteristic of industry that may act as a driver to e-commerce is the existence of strong **national retail networks** such as in Japan. An example is Seven-Eleven's '7dream.com' service in Japan, which allows customers to place orders online then pick them up and make payment at a local store. This indeed plays an important role to enable e-commerce in Japan as a complement to, rather than a substitute for traditional brick and mortar retail outlets.

(b) Partners

Pressures from trading partners are expected to be one of the most critical factors for e-commerce adoption by small firms; as the weaker partners in inter-organizational relationships, small business are extremely susceptible to impositions by their large partners. Such pressures are especially prevalent in the case of e-commerce because of its network nature.

A powerful partner may pursue three different strategies to induce small partner to adopt e-commerce in the relevant way. The first type is recommendation. Powerful partner use information to alter their smaller trading partners' general perceptions of how their organizations might more effectively operate via the use of e-transactions. In contrast, the other two strategies require compliance from the smaller firms. Promises include all tactics that suggest that the larger firm will provide the smaller partner with a specified reward if it becomes e-commerce capable. Threats, on the other hand, refer to actions that convey the larger firm's intentions to apply negative sanction should be the smaller company to become e-commerce capable. Such threats have been evidenced by large automobile manufacturers and department store chains in recent years.

(c) National environment

This part investigates some national characteristics considered as the most important determinants toward e-commerce diffusion of SMEs in various regions around the world. They include: (1) level of government support, (2) level of national ICT infrastructure, and (3) national culture.

Level of Government support

Governments should be a facilitator of business networks among SMEs and between small and large firms. The supports of government for accelerating e-commerce within SMEs are various including training and consultancy, offering state-of-the-art facilities for

events and business transactions, building up Internet homepages for SMEs and other means to improve access to information on business opportunities in foreign markets, helping SMEs participate in foreign exhibitions and trade fairs, prepare marketing materials in foreign languages and circulate advertisement abroad, and above all, the leadership role in accelerating the uptake of e-commerce by SMEs.

The support of the government is reflected, at first, in one of the most complex factors toward SMEs - the technologies associated with e-commerce. Security, electronic identifiers, authentication, digital signatures and passwords and the like are considered to be outside the domain of most SMEs. However, at the stage of adopting e-commerce into their businesses, SMEs needed to become "overnight" experts in the technologies associated with e-commerce. Then, greater co-ordination of training and education activities in IT areas, greater co-ordination of government initiatives in IT areas, and more focused provision of information in an effort to reach more SMEs may give the right incentives for SMEs to involve into e-society.

Level of national ICT infrastructure

Technological change creates new markets and business opportunities. EDI and other new Internet communication tools make cross-border networking easier and more practical for SMEs. They can reduce search costs for potential foreign business partners and improve a firm's visibility in global markets. They drive the internationalization of SMEs in various ways, owing to the growing ease of communications and the high costs of research and development. These tools allow firms in different locations can share know-how, information, distribution networks & other assets simultaneously. Rapid development of ICT has created a more favorable business environment for partnerships and encouraged growth in international strategic alliances. Increase use of ICT by national, regional and local geography, may create an incentive for SMEs to adopt e-commerce. In addition, it may provide a mechanism by which firms could reduce their administrative overheads.

However, the perceived barriers were pointed out in previous researches confirm that SMEs are facing specific problems in fully embracing the opportunities offered by ICT. With the distinguished standards of ICT, Japanese SMEs could not open their trading network toward new potential customers in other countries. Even that communication tools in Japan are powerful above all other advanced technology countries, Japanese SMEs could not diffuse this competitive advantage to more attractive markets overall the world.

Also, discussion in existing researches also figured out that in developing countries, lack of telephone lines, poor accessibility, limited and poor IT support services, poor quality of ISPs are key factors preventing SMEs

from uptake e-commerce as one competitive distribution channel.

Cultural differences

It is commonly agreed that culture encompasses values, beliefs, practices, and traditions that unify a group of people. Therefore, culture is expected to strongly influence one's attitudes toward e-commerce adoption.

At first, cultural factors take greatly effects on e-channel design regarding to the intention to set up a novel distribution e-channel. Also, the influence of culture on marketing has been observed and discussed in a great number of previous studies. For the countries who are conservative like Japan, culture may become one of the highest barriers blocking SMEs in trading domestic instead of internationalization.

Looking on another issue - select trading partners, firms have intention to be based the decision on various decision variables, which could include: partner characteristics; bid characteristics, and the like. Firms placed varying degrees of importance on these variables, depending upon the actual situation. However, standing as socio-economic entity instead of just a pure economic one, Vietnamese SMEs set up its partnership alliances based on individual relationships. Long term association between the parties also provides a sense of community and enhances social bonding. This plays an important role especially for e-commerce diffusion wave.

2.4 Role of SME's leadership toward the adoption of e-commerce

Standing as enterprises with few persons involving, SMEs are lay on the decision and behavior of the leadership in every manner, even this is concerning with financial, human resource, or a technological issue. Then, it is not surprising that for case of SMEs, the diffusion ability of technological innovation in general and Internet in particular lay on the behavior of the leadership in deed.

Openness to new ideas, freedom from investigation of operation, and risk-taking attitude may lead the leadership toward the adopting e-commerce and taking challenges coming with this new innovation. It is well observed evidences of SMEs' leaderships with innovative attitude accelerating their company to engage e-commerce. Lal (2002) indicated that the leadership of this type even taught themselves the know-how of web designs and built up commercial websites without any supports from a professional service provider. Whereas, risk-adverse attitude, incompetence in managerial and technical skills concerning with ICT may prevent the leadership of SMEs from uptake of e-commerce, generate the lack of confidence concerning e-transaction, and then lead them to the decision of not taking the new chance of engaging virtual markets.

3. IMPLICATIONS OF THE MODEL OF DISTRIBUTION E-COMMERCE CHANNEL ADOPTION WITHIN SMES

The above discussed factors frequently indicated in the literature are not impact SMEs' e-commerce adoption decision individually but integrated in a four component model. Interestingly, this model does not just process in term of e-commerce adoption decision only, but also analyze determinants of designing digitalized distribution e-channel as well. This model proposes that new distribution channel with the support of ICT results as an interaction among four components: the collectively perceived benefits, the organizational readiness, the considerations on external environments and attitudes of leaderships.

One important omission in all of the current literature is the lack of real in-depth analysis of the contribution of the external environment component to the e-business venture of SMEs. Most studies are concentrated on the internal factors and tried to point out the different in e-commerce performing of SMEs in different locations. The lack of understanding the role of external factors such as government support, existing of appropriated partners, or cultural characteristics would lead SMEs to be lack of confidence in implementing e-transactions and consequently, hesitate in making a decision of adopting e-commerce.

In fact, due to always in lack of capital and qualified human resources, SMEs needed some general platform to facilitate not only their ICT implementation but also their e-transaction and e-partner finding process that may be facilitated by regional and local governments. The model proposed within the paper (figure 1) suggests that the support of government in building up general ICT infrastructure, general electronic payment system, general distribution systems and the like is a key component in the innovation process for SMEs, in both advanced technology countries and developing countries as well.

The other striking point is that the significant of the leadership characteristics and competences is not get the right attention in literature reviewed to-date. Very little existing researches (Lal, 2002), whereas discussing the role of leadership, just bounded in indicating the effort of SMEs' leadership in getting familiar with new ICT-related know-how and knowledge.

The model proposed in figure 1 suggests that characteristics and competences of leadership is a key component in the e-commerce adoption process for case of SMEs. This means the complex interaction of other three components may not accelerate SMEs to engage e-commerce if leadership of SMEs is conservative and risk-adverse. These attitudes of the leaderships may lead SMEs not to recognize new prospectus generated by the digitalized global network.

However, the support of government or the participating in e-alliance solely does not able to accelerate SMEs to engage e-commerce. The integrate of external environment with all other factors such as organizational readiness, perceived benefit, and especially, the innovative attitude of entrepreneur will compose the powerful to accelerate SMEs in participating in virtual societies and answering the need of ongoing flexibility in the new digital age.

In short, the model demonstrated in figure 1 not just discovered critical factors separately but also reviewed them within comprehensive interactions. This enables us to better understand the nature and the importance of the individual contribution and interaction of the various model components. Studying these interactions would be considered as a starting point to identify how to select a right e-channel model for case of SMEs.

4. CONCLUSIONS

By exploring the literature, this work has identified a model of determinants of e-commerce channel adoption and performance with four major groups namely perceived benefits, organizational readiness, external environment, and attitudes of leadership. The analysis gave an overview of the engine bringing new business values SMEs need to compete, and to produce sustainable revenues in the years to come. The paper also contributes a new way to look at e-commerce innovation within SMEs and reflects a need to better perceive the detailed interaction of the four components of the model.

With the analytical framework composed within the paper, it is supposed that further studies may be taken to generate e-commerce scenarios applicable for case of SMEs. Further insights on each determinant may help to eliminate the tough barriers of supply-side gaps or demand-side gaps toward the uptake of e-commerce by SMEs and suggests the effective ways to gain much more benefit with help of e-commerce for case of SMEs. The framework may also be applied to further discover impacts of e-commerce on distribution channel performance and then to find out the mystery of success in adopting e-commerce as a domestic distribution channel vs. an international distribution channel.

ACKNOWLEDGEMENT

The author would greatly appreciate Professors Eisaku Otsuka and Yoshiki Matsui (Yokohama National University, Japan) for invaluable support and guidance in writing this paper. I wish express my gratitude also to the Telecommunications Advancement Foundation for supporting my work.

REFERENCES

- [1] Akaah, I.P., Korgaonkar, P.K., and Lund, D., "Direct marketing attitudes", *Journal of Business Research*, Vol. 34, No. 3, pp. 211-219, 1995
- [2] Damanpour, F., "The adoption of technological, administrative, and ancillary innovations: Impact of organizational factors", *Journal of Management*, Vol.13, No. 4, pp.675-688, 1987
- [3] Doherty, N., Ellis-Chadwick, F., and Hart, C., "An analysis of the factors affecting the adoption of the Internet in the UK retail sector", *Journal of Business Research*, Vol.56, No.11, pp.887-897, 2003
- [4] Drew, S., "Strategic uses of e-commerce by SMEs in the East of England", *European Management Journal*, Vol.21, No.1, pp.79-88, 2003
- [5] Grewal, D., Iyer, G.R., and Levy, M., "Internet retailing: enablers, limiters and market consequences", *Journal of Business Research*, Vol.57, No.7, pp.703-713, 2004
- [6] Gruber, H., "Competition and innovation: The diffusion of mobile telecommunications in Central and Eastern Europe", *Information Economics and Policy*, Vol.13, No.1, pp.19-34, 2001
- [7] Kendall, J.D., Tung, L.L., Chua, K.H., Ng, C.H.D., and Tan, S.M., "Receptivity of Singapore's SMEs to electronic commerce adoption", *The Journal of Strategic Information Systems*, Vol.10, No.3, pp.223-242, 2001
- [8] Kimberly, J.R., and Evanisko, M.J., "Organizational innovation: The influence of individual, organizational, and contextual factors on hospital adoption of technological and administrative innovations", *Academy of Management Journal*, Vol.24, No.4, pp. 689-713, 1981
- [9] Lal, K., "E-business and manufacturing sector: a study of small and medium-sized enterprises in India", *Research Policy*, Vol. 31, pp.1199-1211, 2002
- [10] Levis, R. and Cockrill, A., "Going global – remaining local: the impact of e-commerce on small retail firms in Wales", *International Journal of Information Management*, Vol.22, pp.195-209, 2002
- [11] Levy, M. and Powell, P., "Information systems strategy for small and medium sized enterprises: an organizational perspective", *The Journal of Strategic Information Systems*, V. 9, No.1, pp.63-84, 2000
- [12] Mehrtens, J., Cragg, P.B., and Mills, A.M., "A model of Internet adoption by SMEs", *Information and Management*, Vol. 39, pp.165-176, 2001
- [13] Meyer, A.D., and Goes, J.B., "Organizational assimilation of innovations: A multilevel contextual analysis", *Academy of Management Journal*, Vol.31, No.4, pp.897-923, 1988
- [14] Olson, J.R., and Boyer, K.K., "Factors influencing the utilization of Internet purchasing in small organizations", *Journal of Operation Management*, Vol.21, pp.225-245, 2003
- [15] Pham, T.T.H., "E-commerce as a novel distribution channel for SMEs: An Exploratory of Japanese SMEs", <http://www2.igss.ynu.ac.jp/~d02lc902/document/NovelChannel.html>, 2003
- [16] Poon, S. and Swatman, P., "An exploratory study of small business Internet commerce issues", *Information and Management*, Vol.35, No.1, pp.9-18, 1999
- [17] Quayle, M., "E-commerce: the challenge for US SMEs in the twenty first century", *International Journal of Operation and Production Management*, Vol.22, No.10, pp.1148-1161, 2002
- [18] Tiessen, J.H., Wright, R.W., and Turner, I., "A model of e-commerce use by internationalizing SMEs", *Journal of International Management*, Vol.7, pp.211-233, 2001
- [19] Van den Poel, D. and Leunis, J., "Consumer Acceptance of the Internet as a Channel of Distribution", *Journal of Business Research*, Vol.45, No.3, pp. 249-256, 1999
- [20] Zheng, J., Caldwell, N., Harland, C., Powell, P., Woerndl, M., and Xu, S., "Small firms and e-business: cautiousness, contingency and cost-benefit", *Journal of Purchasing and Supply Management*, Vol.10, No.1, pp.27-39, 2004

E-business Management System Based on Coordinated Center for Dealer

Ding Zhang¹, Jingwen An¹, Zhongqing Cheng²

¹ School of Management, China University of Mining & Technology (Beijing), Beijing 100083, China

² Department of Logistic Command & Engineering, Naval University of Engineering, Tianjin 300450, China
bace@tom.com

ABSTRACT

This management platform has been applied in e-business of the dealer. Based on coordinated center, and it is realized by intranet, extranet and Internet. The structure of the management platform has five parts: information center of management platform, customer relationship management system, purchasing management system, logistics management system and financial management system. With network security, data security, user security and backup etc., the system reliability and security are enhanced. Through Business to Business (B2B) and Business to Customers (B2C), it can help the dealer coordinate with the supplier and the retailer.

Keywords : dealer, e-business, management system, security, B2B, B2C

1. INTRODUCTION

Along with the development of market competition and network economy, traditional business has been transforming to e-business, and requiring higher and higher speed and accuracy of information much more than before. As a role in the marketplace, the dealer needs more and more information to coordinate with the supplier and the retailer. It has been an important problem for the dealer to improve management level and to fulfill the requirements of customers, the supplier and the retailer. Nowadays, the e-business of the supplier and the retailer has grown rapidly. And then, that makes all dealer's act depend on e-business, such as logistics, accounting and information etc. E-business is the development trend of management. By taking commercial activities as e-business system, the dealer can strengthen the competition capability in future. And now the cost of software and hardware is reduced, so the dealer need not pay large quantity of investment for e-business.

For the purpose of e-business, this system is built as a platform and coordinated center for the dealer. The system has the functions of B2B and B2C electronic business system, and can be remote controled through intranet, extranet and internet, connecting the supplier, the retailer and the dealer themselves. It is also realized by circle movement of logistics, information flow and fund in effect: logistics is by sale services and delivery, information flow is by Electronic Data Interchange (EDI) and Web Page Browsing, fund circulation is by electronic payment. Though this system, information feedback is in time. It can keep sensitive with the changes of external environment and relations from environment analysis, and can connect other enterprise, for example, the third party logistics. With this system, the dealer can be an excellent dealer in the era of e-business, which has competitive advantage, scales, coordination ability and management capability.

2. BUSINESS SYSTEM MODEL OF DEALER

The dealer has other titles, which to be used by people: middleman, broker or agent. Its responsibilities are keeping the cooperation relationship with the supplier and the retailer. It takes charge managing logistics, finance, inventory and customer satisfaction. It coordinates with the supplier, the retailer and customers, and exchanges information with the supplier and the retailer. The relation framework is illustrated in figure 1.

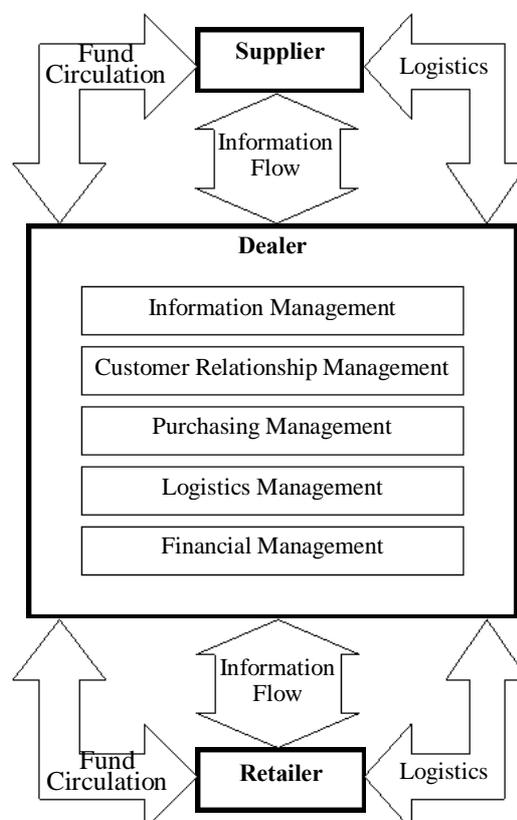


Figure 1. Business system model of dealer

3. STURCTURE AND FUNCTION OF THE E-BUSINESS SYSTEM

Based on supply chain of B2B, e-business system is constitutive of the dealer, the supplier and the retailer, which exchanges information as a main platform. The system also has a function of B2C. Depending on large commodity database, the system can make directory of commodity, OnLine Analytical Processing (OLAP) order, give real time report of transaction and commodity inventory, publish commodity information and price. And it can support commodity searching in many aspects, online marketing, and have automatic analysis system with information from the supplier and the retailer, give service for customers' requirement and suggestion, query executive information of order, and have ability of safe payment .

The system has five parts, the main framework is illustrated in figure 2.

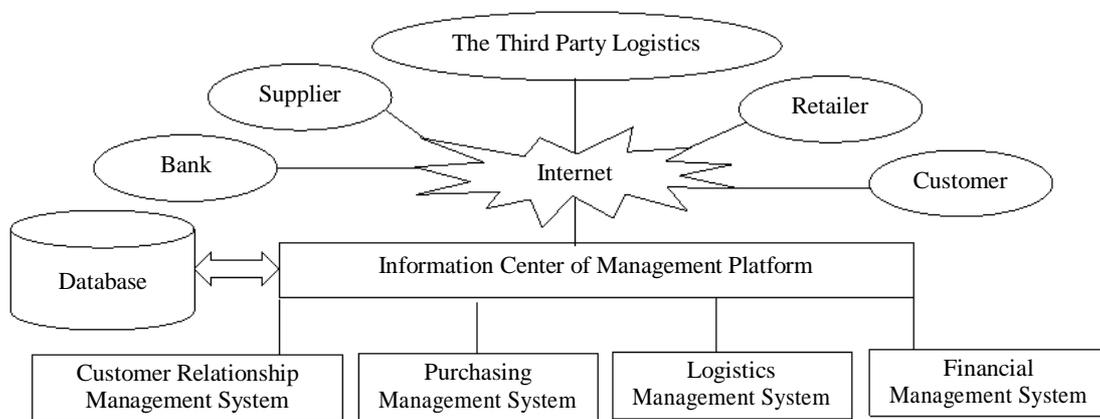


Figure 2. Framework of the e-business system

3.1 Information Center of Management Platform

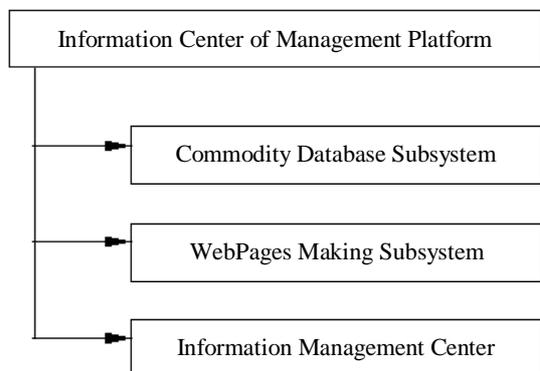


Figure 3. Structure of information center of management platform

Information center is one of the ways to integrate supply chain in Internet. With data storage and information

processing, admitting to download key data of supply and demand, the center can process and transfer information to all participants continuously. The whole network format a Hub-and-spoke system. And participant's information system in intranet is the auxiliary system of this center. For example: ERP and other enterprise's system. The concrete structure are as followed (illustrated in figure3).

3.1.1 Commodity database subsystem

Base on the different types of commodity and the supplier, the subsystem makes classification and encoding of commodity, and inputs the information into database, which includes the supplier information, commodity introduction, commodity specifications, pictures, FAQ and price etc.. This can be completed by personnel, and can change the information of commodity in any time.

3.1.2 WebPages making subsystem

This subsystem takes charge to make the WebPages of the dealer and puts into Internet. It has two subsystem.

Manual making subsystem: The personnel designs WebPages with software. It always applies in homepage and other irregular pages.

Automatic generating subsystem: This subsystem can automatically create specifications WebPages in fixed modes. Because the dealer need make introduction page for all commodity, and the dealer has much more types of commodity, it will be easy to create WebPages by the subsystem.

3.1.3 Information management center

This center manages whole intranet and extranet to assure the information flow unblocked. It manages operation and maintenance of network and Website, supports net connection and management software.

Information publishing, maintenance and management:

The publication of news and information are automatic. This center can manage the information ,and supply users to query.

Notice information transfer: This subsystem can send notice of market by message or mail. It can send news and e-mail from Mail Server, send the important business information, such as customers' order and information feedback, and can connect with mobile phone in time, in case to miss market opportunity.

3.2 Customer Relationship Management (CRM)

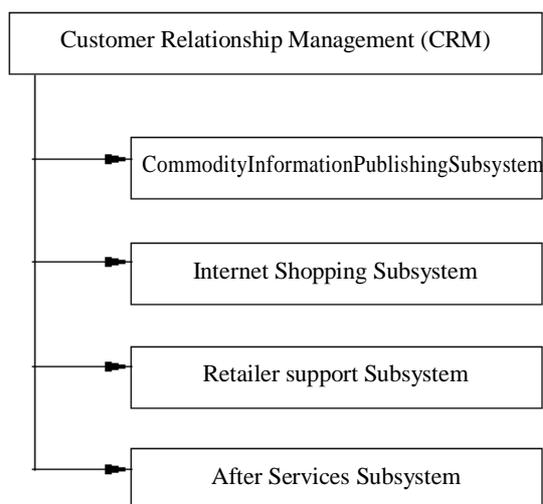


Figure 4. Structure of customer relationship management system

This system has two functions, sale management and customer support, and has two customer classification, one is customer and the other is retailer. It will realize Networked marketing and customer service, and provide customer analysis, consumer behavior tracking and One-to-One marketing strategy. The sale management can arrange work calendar and make up schedule, realize relation management, sale forecast, make suggestions, pricing management, distributive of market and cost report etc.. With this system, we can improve efficiency and reduce costs. The customer support can manage the customers' order, provide customer query, commodity directory and technical support. It also can provide online marketing guide to customer and retailer. It includes tracking order, customers concern, dispute solution and service management. There is a database to record problems and solution. It can make repairing, arrange scheduling, assist personnel to serve customers,solve problem quickly, and report deep market information.The structure of the system is illustrated in figure 4.

3.2.1 Commodity information publishing subsystem

Customers can search the information and inventory of

commodity by two ways, classifying directory and keywords.

They can search the commodity which they need with the type and manufacturer.Customers can fast inquire with keyword, which is brand name, manufacturer name or others, and can be more than one.

3.2.2 Internet shopping subsystem

The trade flow is a virtual supermarket. When customers are viewing commodity, they can put commodity into Internet shopping bag or replace to shelf. And they can check their bag, change commodity sums, settlement, choose payment mode and deliver mode in any time. After customers make an order, subsystem will send order message to customers and salespeople, and then update the content of order management system. Customers can query the execution of order.

3.2.3 Retailer support subsystem

Every retailer has its own account. They can query price, support decision, market mode, information of new products and payment. They can send repairing message to the dealer and check the cost. And based on the user name and commodity code which customers sent to the dealer before, the dealer will check the information and give the cost.

3.2.4 After services subsystem

The dealer can send information and answer questions to consumer, make commodity maintenance and repairing, or withdrawal.

3.3 Purchasing Management System.

This system has two subsystem: the supplier query subsystem, and order management subsystem, The structure is illustrated in figure 5.

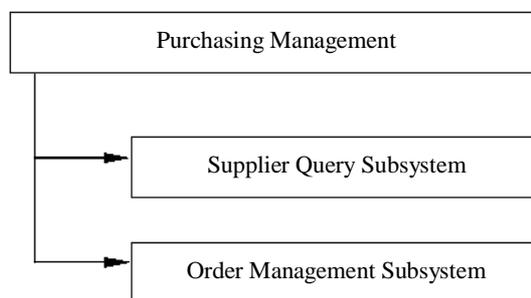


Figure 5. Structure of purchasing management system

3.3.1 Supplier query subsystem

Through Internet, the supplier can query status of sales, accounting and inventory. So they can provide

commodity in time, and set good relations with the dealer.

3.3.2 Order management subsystem

This subsystem can make classification encoding of the supplier, the retailer, storage. And it can order and arrange dispatching according to customers' order. Purchasing department can communicate with financial department with the subsystem, practise purchase strategy, control purchase cost and design purchase process rationalization. So it can reduce cost, shorten purchase cycle and improve quality of purchase.

3.4 Logistics Management System

This system is one of keys for the dealer to realize MIS (Management information System). It is the basis of logistics management, which includes operation subsystem and inventory management subsystem. Assisted by Purchase-Sales-Inventory Management System, the dealer can control content of inventory and market, reduce costs of purchasing and logistics, increase turnover rate, optimize resource reorganizing, analyze marketing environment and get sale fluctuating, so to improve itself competition capability. The structure is illustrated in figure 6.

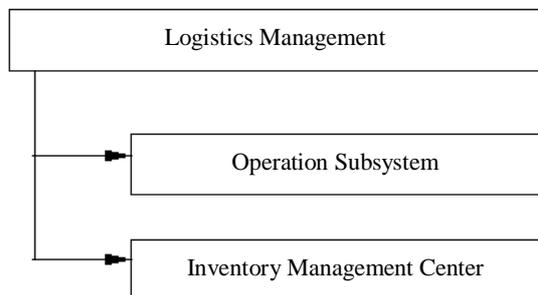


Figure 6. Structure of logistics management system

3.4.1 Operation subsystem

Logistics delivery is the main task of operation subsystem. After customers send order, the subsystem will automatically send order message to the dealer. According to the order and inventory, the subsystem can arrange logistics delivery and vehicle routing to satisfy different requirement of customers

3.4.2 Inventory management center

Inventory management includes management of commodity passed storage in and out, alarm for overstock and overtime, and purchase from the supplier. According to executive case of order, operation personnel can update the content of order to management system at any moment. The subsystem can support dynamic search with content of order management system and inventory, to help manager

understand the content of sales and inventory. Bass on inventory minimum limit, subsystem can automatic alarm, apply for purchase.

3.5 Financial Management

The system embodies stratagem management, business management and control so and so functions of accountant. It includes three parts as followed. The framework is illustrated in figure 7.

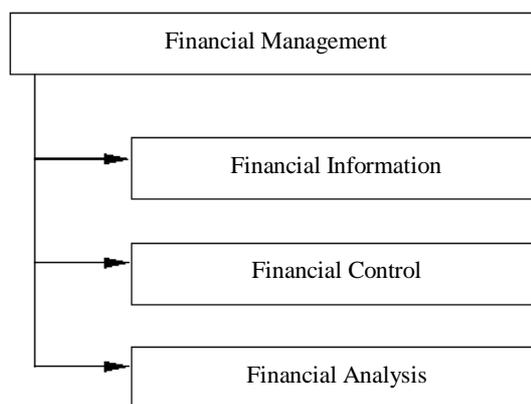


Figure 7. Framework of financial management system

3.5.1 Financial information

The enterprise need collecting and save the information, and should assure the real-time and relevant, reliable and comparable, in order to keep the operation run normally.

3.5.2 Financial control

Financial control can guarantee the validity of management, so as to improve the sale level, decrease expenses and realize the purpose of promoting the economy benefit of the enterprise. This system can manage accounting and realize reporting. It support electronic payment system, with which, customers can enter net-bank and use Secure Electronic Transaction Protocol or Security Socket Layer. After they input account number and password, they can query their accounting and payment. This system can realize user authentication, accounting and user management. The merit of this system are real-time performance, powerful security and RAS (Reliability Availability and Service).

3.5.3 Financial analysis

The dealer need analyze finance and market information by analytical method and evaluation index system, which serve as a channel for managers to master financial status and evaluate the profit condition, for the financial state analysis of the typical state owned enterprises.

4.METHOD DEFENDING SECURITY OF E-BUSINESS SYSTEM

As an e-business platform, this system needs good security to assure the reliability.

4.1 Network Security

With data encryption, firewall and access control, the subsystem keeps network security well, and with RAID, Error Detection and Correction (EDAC) and software fault tolerance etc., to assure system reliability.

4.2 Data Transmission Security

This subsystem can use instant messaging and electronic data interchange to change the electronic data from SSL, RSA and RES etc. to assure data security. And classified information is encrypted, before it is translated through Intranet.

4.3 User Security

All the management and operation process are realized by the browser. And the privileges grading control is applied in the system. The only administrator has the highest-grade permission, who can change the username and password, but can not add and apply other administrator. When administrator accesses, he is not like an ordinary user, he can add other user and authorize grade permission. For example, he can add a specialty user for financial manager, who can only query financial information. And he can add another specialty account for salesperson and give corresponding permission. The user security can be assured by user password, mixed cryptogram and data encryption etc.. Because the permission of administrator is in the highest-grade, it is mainly used in setting and managing whole virtual market. Considering the system security, it should be better not to use the user frequently.

4.4 Security, Data Backup, Log Recovery, Restore and Monitor

This subsystem can set the time interval of data backup, or backup immediately. And the manager can check log file and running state.

5.CONCLUSIONS

Facing e-business and enterprise online, this system set up a platform for the dealer as a coordinated center. Based on the conception of Open, Cooperative and Services, union management of resource, with

e-business platform as background, the system builds a main supply chain with the dealer, the supplier and the retailer, which has transaction function of B2B. As an e-business platform, it can communicate with Website from Internet, and provide online marketing to customers as B2C. To whole management system, software adopts standard interface to connect the mainstream financial software and enterprise website. And it can connect with other e-business website, such as CRM, ERP and MIS etc. So it is adaptable. This system has good security, perfect service, standard common interface, friendly user interface, mass user capacity, processing ability and complete user management. This system has large commodity database, and can make directory of commodity, publish list and picture of new commodity and hot commodity, OLAP order, give real time report of transaction and commodity inventory, publish price. And it supports commodity searching in many aspects, and has automatic analysis system with information from the supplier and the retailer, gives service to customers' requirement and suggestion, query executive situation of order, and has online payment, all of these are secure. Based on Intranet of Web, the system brings great flexibility. The dealer can coordinate and communicate with the supplier, the retailer and customers with this system, so as to be an indispensable role between the supplier and the retailer.

ACKNOWLEDGEMENT

The authors are grateful for the support of the E-business Management System from Shenzhen Suige Industrial Development Co. Ltd., China.

REFERENCES

- [1] Ash, C.G. "E-Business Change and Organisational Performance: A Case Study of B2B e-Business Integration", *Enterprise Information Systems III*, pp 287-294, 2002. Kluwer publications, Dordrecht, Netherlands /ISBN 1-4020-0563-6 Edited by Joaquim Filpe, Bernadette Sharpe, Paula Miranda
- [2] Diao, Xing-chun., "A Typical Model of Electronic Commerce Platform", *Mini-Micro Systems*, Vol. 24, No. 10, pp1856-1860, 2003.
- [3] Mohamed Alseraji, Chen Song-qiao, "Small I/E Company Automation System Developing and Design", *Computing Technology and Automation*, Vol.22, No.1, pp108-110, 2003.
- [4] Wei Wang, Shan-ping Li, "Framework Research of Building E-business Information Infrastructure", *Proceedings of International Conference on e-Business 2002 Beijing(ICEB2002)*, pp219-222, May 23-26, 2002, Beijing, China.

E-Business Oriented Optimal Online Auction Design

Anshi Xie¹, Yijun Li¹, Yong Li²

¹School of Management, Harbin Institute of Technology, Harbin 150001, China

²Bio-Chemical Engineering College, Beijing Union University, Beijing 100023, China

¹Stanford@hit.edu.cn, ²chestnut_ly@sohu.com

ABSTRACT

Online auctions, in the absence of spatial, temporal and geographic constraints, provide an alternative supply chain channel for the distribution of goods and services. This channel differs from the common posted-price mechanism that is typically used in the retail sector. In consumer-oriented markets, buyers can now experience the thrill of 'winning' a product, potentially at a bargain, as opposed to the typically more tedious notion of 'buying' it. Sellers, on the other hand, have an additional channel to distribute their goods, and the opportunity to liquidate rapidly aging goods at greater than salvage values. The primary facilitator of this phenomenon is the widespread adoption of electronic commerce over an open-source, ubiquitous Internet Protocol (IP) based network. In this paper, we derive an optimal bidding strategy in sequential auctions that incorporates option value assessment. Furthermore, we establish that our optimal bidding strategy is tractable since it is independent of the bidding strategies of other bidders in the current auction and is only dependent on the option value assessment

Keywords: online auction; optimal auction; mechanism design

1 INTRODUCTION

1.1 Auction theory

As an old price-discovering mechanism, auction entered the economics literature relatively recently. Two papers have been regarded as initial references [21, 44]. The full flowering of theory came only at the end of the 1970s. A very readable introduction to the state of this field by the late 1980s is in McAfee and McMillan [32]. Another helpful introductory article is Maskin and Riley [31]. Why should we study auction? McAfee and McMillan explained that auction theory provided one explicit model of price making. A less fundamental but more practical reason for studying auction is that auction is of considerable empirical significance. The value of goods exchanged each year by auction is huge. This fact in itself indicates that some theoretical study of auction is warranted. More over, as will be seen, the theory of auction is closer to applications than is most frontier mathematical economics [20]. Next, we will give an overview on this topic from three perspectives: definition, schemes and economic models.

Definition

Oxford Dictionary defines an auction as a "public sale in which articles are sold to maker of the highest bid". This definition describes the English auction, which cannot reflect the essence of auction. An authoritative definition is given by McAfee and McMillan [32] as follows: an auction is a market institution with an explicit set of rules determining resource allocation and prices on the basis of bids from the market participants.

Basic auction schemes

This section will give the reader a brief overview of the auction schemes.

Although there are various kinds of auction classifications, there appears to be no unified forms. Here we briefly summarized each classification and discussed them from traditional classifications.

a. According to the traditional view, we can divide auctions into "outcry auction" and "sealed-bid auction". Outcry auction includes the ascending-price auction (English Auction) and the descending-price auction (Dutch Auction). Sealed-bid auction includes the first-price sealed-bid auction and the second-price sealed-bid auction. It should be pointed out that it was Vickrey who first defined the second-price sealed-bid auction. As such, the second-price sealed-bid auction is usually also called the Vickrey auction. Jason and Margolis [23] presented an nth price mechanism, which, however, didn't receive any recognition from other researchers.

b. According to the market structure, we divide auctions into single auction and double auction. Single auction is a "one to many" auction in nature. Either buyer or seller monopolizes the market resource and holds the market. Double auction differs from single auction in structure. Double auction is "many to many". Both sides are related through supply and demand. An environment where vendors and buyers meet with the goal to sell and buy goods is commonly called a market. As there are many different interpretations of what a market is, such an environment should rather be called a market framework. Figure 1 shows an example of such a framework. There are four cases presented: 1. One vendor and one buyer directly *negotiate* in the classical sense; 2. Multiple vendors and one buyer are engaged in a *reverse auction*; 3. Multiple buyers and one vendor are engaged in a classical *auction* (English, Dutch, Vickrey,

etc.); 4. Multiple buyers and vendors trade in a *market*.

c. According to the value form, we divide auctions into common-value auction and private-value auction. In certain auction, because of the information asymmetry, only that buyer knows one buyer's valuation of a product; not all buyers may agree on a product's value; buyers may value a product differently for different reasons. We call this auction private value auction. The main commodities for this auction type are paintings, Egyptian thimbles, and Victorian snuffboxes etc. In a common-value auction, the actual value of (or any known measure of) the goods is the same for all buyers. Buyers start with different estimates or private information about the following aspects of the goods: quality of the goods, remaining economic life of the goods, and some "measure" of the goods. In a common-value auction, buyers can receive meaningful information about the product's "true value" from other buyers' bids. The typical commodity for this auction is mineral rights. In a private-value auction, buyers rarely get any information about the product's "true value" from other buyers.

d. According to the status of buyers and sellers, we can divide traditional auctions into forward auction and reverse auction. In reverse auction, sellers bid for a buyers' business. The seller that makes the lowest bid wins the auction. Just as there are open and sealed bid auction in the "regular" auction format—there are open and sealed bid formats in reverse auction too. The new e-procurement system is a typical reverse auction system. Buyers are on a good wicket, and sellers bid for commodity.

1.2 Auction online

Internet allows direct contact between buyers and sellers, and provides a suitable and popular place for conducting auction, which are usually called e-auction or auction online. We define e-auction as a web page, which displays information about a good or service with the intent to sell it through a competitive bidding procedure to the highest bidder. In fact, e-auction is an application environment of auction, so we cannot study e-auction as a separate research field. Most literature on e-auction is about its role in enterprise [2, 18, 24, 48].

There are many differences between traditional auction and e-auction. E-auction gives bidders increased convenience both geographic and temporal. These increased conveniences can also benefit the seller by creating a larger market for auctioned good. Furthermore, search engines and clickable hierarchies of categories for browsing make it more convenient for a bidder to find the goods she is looking for.

The earliest e-auction Web sites are Onsale and eBay which opened in 1995. It was the first to take advantage of the technologies offered by the Web, including the

use of automated bids entered through electronic forms, and search engines and clickable categories to allow bidder to locate their items of interest. In the past several years, two large Web-based companies, Yahoo! and Amazon, have announced their own person to person auction services very much like eBay's. It will be interesting to see whether these well-funded later entrants will be able to take business away from eBay. E-auction is potentially rich sources of study both for economic theorists and for empiricists.

In electronic auction, we can have readily available items and face global market; we can easily access information and have a large scale of operation; we have flexible duration and can compete with other auctions and retailers. The advantages mentioned above indicate that electronic auction will ultimately and inevitably replace traditional auction.

Business-to-consumer online auctions are heating up as an efficient and flexible sales channel. Along with the other two types of online auctions, namely consumer-to-consumer and business to business *auctions*, they represent a new class of mercantile processes that are ushering in the networked economy, but are not fully understood yet. Yong and Yum^[51] have called for an extensive examination of the pervasive impact of advance electronic communication on the theory and practice of auctions. For a wide variety of goods sold over the Internet, consumers now have an interesting choice of mercantile processes to utilize to buy these goods. Broadly, these different processes can be broken into static posted-price mechanisms and dynamic interactive pricing mechanisms such as online auctions. In this context, online auctioneers are striving to find the correct strategic and tactical direction that ensures their long-term profitability. This involves optimizing existing as well as designing interesting new mercantile processes to sell their products. This paper is an attempt to analyze this emerging market structure with respect to the welfare of the different economic agents involved namely the *B2C* auctioneers (we implicitly assume the joint interest of the seller and the auctioneer, if they are different) and consumers. There already exists a vast body of literature dealing with the theory of auctions and optimal mechanism design. However, the majority of the work focuses on the analysis of single auctions and is carried out in isolation from the broader context of the markets in which these auctions take place. It should also be noted that despite the presence of a vast body of literature dealing with traditional auctions, one could not call these auctions 'mainstream' from a consumer's viewpoint. It is only after the synergetic combination of traditional auctions with Internet technology that this sphere of economic activity has blossomed into significance.

In the next section of this paper, we derive an optimal bidding strategy in sequential auctions that incorporates option value assessment. Furthermore, we establish that

our optimal bidding strategy is tractable since it is independent of the bidding strategies of other bidders in the current auction and is only dependent on the option value assessment.

2 OPTIMAL AUCTION DESIGN

What is the best auction design to use for a given selling situation? The work covered below all assumes a risk-neutral seller selling a single indivisible item to one of possibly several potential buyers.

In cases in which the bidder are risk-neutral, symmetric, and follow the IPV model, the Revenue Equivalence Theorem holds (RET). The RET says that under certain conditions, the four major auction types—English, Dutch, First Price Sealed bid, and Vickrey auction will deliver the seller the same expected revenue. In this case, the optimal auction requires the seller to set a reserve price below which he will not sell the item. This price is set to mimic the expected bid of the second-highest bidder, and is strictly greater than the seller's reserve valuation. Moreover, the seller should go ahead and announce this reserve price before the auction begins.

Weakening the assumptions of the RET leads to changes in the optimal auction design. In cases where the bidders are risk-averse, but the assumptions of symmetry and IPV still hold, the Dutch and the First price sealed bid auctions. A seller should still optimal set a reserve price and possibly add an entry fee as well.

In cases where the bidders have affiliated values but remain risk-neutral and symmetric, sellers should use the English auction because it yields the highest expected revenue, followed by the Vickrey auction, which is then followed by the Vickrey auction, which is then followed by the last-place tie between the Dutch and First price sealed bid. Moreover, a seller should still set reserve prices and entry fees; in fact, given a choice between two pairs which will attract the same set of bidders, it pays to choose the pair with higher entry fee and the lower reserve price. Side payments from the auctioneer to the bidders are sometimes optimal, but rarely used in practice.

Finally, when bidders are asymmetric, but remain risk-neutral and have IPV, the optimal auction format is highly variable.

The key is that auction mechanism design is theoretically well-grounded. There are different auctions are known to be optimal in different situations. Moreover, auctions for which truth-telling is the dominant strategy do exist and are quite feasible. Hence, an agent's best possible course of action could be carried out by a dumb agent simply telling the truth. There are two important issues in the problem of optimal auction design. One is strategy. The other is protocol. Generally, researchers are focus on one issue.

In the next section, we mainly discuss optimal bidding strategy in different condition.

3 STUDY OF OPTIMAL BIDDING STRATEGY

In this section, we consider a sale of two identical objects through second-price sealed-bid auctions. If a bidder wins an auction, she will leave the game. If she loses the first auction, she will bid in another auction with probability of 1. Each bidder intends to buy at most a single object i.e., the utility of a second object is assumed to be zero. In future work, we intend to relax this assumption. However, we conjecture that this will not qualitatively change the results obtained here as long as marginal utility is diminishing i.e. the utility gained from the second object is smaller than that from the first. In this paper, we define *new bidders* to be those that have not bid in any auctions selling the same item before. *Seasoned bidders* are those for whom the current auction is their second auction. It is easy to understand that a new bidder's expected utility comes from both auctions. The first component of a bidder's utility is the difference between her reservation value and price in the first auction; the second component is the corresponding difference in the second auction multiplied by the probability of losing in the first auction, because only if she loses in the first auction will she receive utility in the second auction. Since we use bidder's expected payoff as her utility function, we are assuming risk-neutrality here. We assume that U_1 is the utility from auction, r is reservation value, U_2 is the utility from second auction and P is the probability of losing first auction, then,

$$U_1 = \begin{cases} 0, & \text{if she loses;} \\ |r, & \text{if she wins} \end{cases}$$

$$U_2 = \begin{cases} 0, & \text{if she loses} \\ |r, & \text{if she wins} \end{cases}$$

In the rest part of this paper, we will assume independent, private valuation, as in the private value model. We will also assume that within the same auction, there are two types of bidders—new bidders and seasoned bidders. We assume, mostly for ease of notation, that bidders of the same type adopt the same type of bidding function (strategy). From private-value model, we know that seasoned bidders would bid their reservation values. Hence, we take the following assumption: New bidders adopt the same type of bidding strategy $b(v)$, where v is a bidder's reservation value. Bidders' valuations are drawn independently from a distribution $F(\cdot)$.

4 BIDDING STRATEGY UNDER CERTAINITY

We start by assuming that bidders know the number of new and seasoned bidders that will bid in each auction. As in all private value models, the bidder is assumed to know the distribution of reservation values of other bidders but the actual value is private information to each bidder. A seasoned bidder A will bid his

reservation value, since for him there are no future auctions. On the other hand, a new bidder, B has to consider how her first auction bid affects the likelihood of being able to participate in a second auction. Before formally deriving B 's optimal strategy, a simple example provides some intuition for B 's strategy.

Example 1: B bids in two auctions. In either auction, B competes with another bidder whose bid is drawn from a uniform distribution $U(0,6)$. B 's reservation value is 4. We consider the following alternatives:

Alternative 1: B bids her reservation value in both auctions. Recall that since this is a second price auction, B wins at its opponents bid x when $x \leq 4$.

$$E(u) = 1.78$$

Alternative 2: B bids less than her reservation value in the first period and reservation value in the second period. If she bids 2 in the first auction, her expected utility is:

$$E(u) = 1.89$$

The increments in her bid will increase her expected utility initially. Once it reaches 2.67, her expected utility begins to drop. Simple calculations show that the optimal bid for bidder B in the first period is 2.67.

We define B 's expected utility if she bids x given her reservation value of v as

$$E(u(x, v)) = \int_0^x (v-z)d\Phi(z) + \int_0^v (1-\Phi(x))(v-z)d\Phi(x) \tag{4-1}$$

The first integral is the bidder's expected payoff in her first auction, where $\Phi(z) = \Pr(b(x^{(1)}) < z \& y^{(1)} < z)$, and where $b(x)$ is the other new bidders' bidding strategy, and $x^{(1)}$ is the largest reservation value among new bidders (excluding bidder B) in auction 1. Hence the largest bid of new bidders is $b(x^{(1)})$. Similarly, $y^{(1)}$ denotes the largest reservation value, and hence, also the highest bid among seasoned bidders in auction 1. Similarly, we define $\underline{x}^{(1)}, \underline{y}^{(1)}$ for the second auction as the highest bids among the new bidders and seasoned bidders (excluding B). Therefore $\Phi(x) = \Pr(b(x^{(1)}) < x \& \underline{y}^{(1)} < x)$ is the probability of B winning her first auction. The second term in equation (4-1) is B 's expected payoff in her second auction.

Solving the first order condition for an interior optimum with respect to x , we have the following optimal first period bid x^* :

$$x^*(v) = v - \int_0^v \Phi(z) dz \tag{4-2}$$

where $\int_0^v \Phi(z) dz$ is bidder B 's expected payoff if she bids in her second auction.

From the above equation, we may derive the following properties of bidder B 's optimal strategy.

Property 1: x^* is B 's dominant strategy in the sense that it is independent of the strategies of other bidders in the first auction.

$\Phi(x)$ includes all the information about the remaining bidders in the first auction. In B 's optimal strategy, there is no $\Phi(x)$, meaning that her first period bid is independent of the strategies of all other bidders. The conclusion may seem surprising but it is easy to understand. If we let v_2 represent B 's expected payoff from period 2 auction, B 's "real" valuation of the first auction is $v - v_2$. Hence, according to the private-value model, B should bid her "real" valuation.

Property 2: $x^* < v$ This property explains why some experienced bidders would like to bid less than their reservation values in the first auction. Simply put, the option of being able to bid in a second auction is valuable. The smaller the value of this option, the greater is the probability of winning in the first auction. Thus the first period bid must trade-off the payoff from the first auction against the option value.

Property 3: x^* increases when N or M increases. Here N_2 and M_2 denote the number of new bidders and seasoned bidders respectively. Intuitively, when the number of bidders increases, bidder B faces more competition in the second auction, which makes the second auction less valuable to her. Hence, B increases her bid in the first auction.

5 CONCLUSIONS

Online sequential auctions are different from non-Internet sequential auctions. In most theoretical models of sequential auctions, the number of auctions is fixed; bidders enter and leave the auctions at the same point (or winners leave and the rest remain). So, the k th auction of bidder A is also assumed to be the k th auction of bidder B . However, in online sequential auctions, no such symmetry can be assumed. Each bidder faces a continuous stream of auctions. Bidders enter those auctions at different times and may have participated in different number of auctions. Thus, the principal contribution of this paper is a model of online sequential auctions that captures important features of these auctions neglected in earlier models of sequential auctions. Next, we will mainly study on how uncertainty affects bidding strategies

6 REFERENCES

- [1] Abrache, J., T. G. Crainic, M. Gendreau. Design issues for combinatorial auctions. *Proceedings of the Fourth International Conference on Electronic Commerce Research*, Dallas. 2002.
- [2] Alan, S., H. Alan. Online reverse auctions and their role in buyer-supplier relationships. *Journal of Purchasing & Supply Management* 9 257-268. 2003.
- [3] Ba, S.I., A. B. Whinston, H. Zhang . Building trust in online auction markets through an economic incentive mechanism. *Decision Support System* 35 273-286. 2003.
- [4] Bass, S. M., H. Kwakernaak. Rating and Ranking of Multiple Aspect Alternative Using Fuzzy Sets. *Automatica* 13 47-58. 1977.
- [5] Ben-ameur, H., B. Chaib-draa. Multi-item Auctions for Automatic Negotiation. *CIRANO Working Papers* 68. 2002.
- [6] Bichler, M. Bidding languages and winner determination in multi-attribute auctions. *IBM Research Report RC 22478*. 1998.
- [7] Bichler, M. Winner Determination Algorithms for Electronic Auctions: A Framework Design . *Proceedings of the Third International Conference on E-Commerce and Web Technologies* 37-46. 1997.
- [8] Bichler, M., M. Kaukal, A. Segev. Multi-attribute auctions for electronic procurement. In *Proceedings of the First IBM IAC Workshop on Internet Based Negotiation Technologies* 18-19. 1999.
- [9] Bichler, M. An experimental analysis of multi-attribute auctions. *Decision Support System* 29(3) 249 – 268. 2000.
- [10] Bleeker, A., L. Meertens A Semantics for BAN Logic. *DIMACS Workshop on Design and Formal Verification of Security Protocols*. 1997.
- [11] Branco, F. The design of multidimensional auctions. *RAND Journal of Economics* 28(1) 63-81. 1997.
- [12] Burmetister, B. A practical Approach to Multi-attribute Auctions. *DEXA Workshops* 670-674. 2002.
- [13] Carr, R. I., J. W. Sandahl. General Bidding Model. *Journal of the Construction Division. ASCE* 108(4) 639-651. 1982.
- [14] Che, Y. K. Design Competition Through Multidimensional Auctions. *RAND Journal of Economics* 24(4) 68-680. 1993.
- [15] Chen, C.J., B.W. Lin. A resource-based view of IT outsourcing: knowledge sharing, communication, and coupling quality. *Asia Pacific Management Review* 6(2) 149-173. 2001.
- [16] Chen, S.Y. Theory and Application of Fuzzy System Decision Making. DLUT Press, Dalian China. 1994.
- [17] David, E., S. Kraus. Protocols and Strategies of Automated Multi-Attribute auctions. *AAMAS 02' conference* No 2531. 2002.
- [18] Dave, C. Explorations in evolutionary design of online auction market mechanisms. *Electronic Commerce Research and Applications* 2 162-175. 2003.
- [19] Dekrajangpetch, S. Interior-point Linear Programming Algorithm for Auction Methods. *IEEE Transactions on power systems* 15(7) 572-578. 2001.
- [20] Feng, W. J. Relative Efficiency in Multiobjective Optimization and Decision Making. *System Engineering and Electronic Technology*. 4(3) 76-82. 1994.
- [21] Friedman, L. A Competitive Bidding Strategy. *Operations Research* 4 106-112. 1956.
- [22] Hunsberger, L., J. Barbara. A Combinatorial Auction for Collaborative Planning. *Proceedings of the Fourth International Conference on Multi-Agent Systems* 151-158. 2000.
- [23] Jason, F. S., M. Margolis. A random nth-price auction. *Journal of economic behavior & organization* 46 409-421. 2001.
- [24] Lee, H.G., T.H. Clark. Impact of the Electronic Marketplace on Transaction Cost and Market Structure. *International Journal of Electronic Commerce* 1(1) 127-149. 1996.
- [25] Lengweiler, Y. The multiple unit auction with variable supply. *Economic Theory Journal* 14 (2) 373-392. 1999.
- [26] Li, Y. J., Y. Z. Cao. The new development of NSS. *Asia Pacific Management Review* 6(2) 197-215. 2002.
- [27] Lin, R., Y. Z. Zhang. Theoretic Analysis on Second Sealed Auction. *Zhejiang Social Science* 3(3) 78-95. 1999.
- [28] Liu, S. L., Research on multi-attribute decision making theory, method and applications. *Paper dissertation*. Beihang University, Beijing. 1997.
- [29] Liu, S. L., S. Y. Wang. Decision Theoretic Approaches to Bidding and Auctions. *Studies in International Technology and Economy* 1(2) 20-33. 1998.
- [30] Maes, P., R.H Guttman, A. Moukas. Agents that Buy and Sell. *Communications of the ACM* 42 81-91. 1999.
- [31] Maskin, E., J. Riley. Optimal auctions with risk-averse buyers. *Econometrica* 52(6) 1473-1518. 1984.
- [32] McAfee, R. P., J. McMillan. Auctions and Bidding. *Journal of Economic Literature* 25(2) 699-738. 1987.
- [33] Milgrom, P., R. J. Weber. A theory of auctions and competitive bidding. *Econometrica* 50 1089-1122. 1982.
- [34] Mishra, D., D. Veeranmant. A multi-attribute Reverse Auction for Outsourcing. *Proceedings of the 13th International Workshop on Database and Expert Systems Application* 675-679. 2002.
- [35] Morin, T. L., R. H. Clough. Competitive Bidding

- models. *Journal of the Construction Division. ASCE* **95**(1) 85-107. 1959.
- [36] Qi, Y.F. Multi-criteria Decision Making. Weapon Industry Press, Beijing. 1989.
- [37] Richard, E. W. On a possible Benefit to Bid Takers from Using Multi-Stage Auctions. *Management Science* **34**(9) 1109-1120. 1998.
- [38] Richter, C. W., G. B. Sheble. Genetic Algorithm Evolution of Utility Bidding Strategies for the Competitive Marketplace. *IEEE Transactions on Power Systems* **13**(1) 256-261. 1998.
- [39] Sandholm, T. eMediator: A Next Generation Electronic Commerce Server. *AAAI Workshop on AI in Electronic Commerce* 46-55. 1999.
- [40] Srividhya, S. Design and Verification of A Secure Electronic Auction Protocol. *The 17th IEEE Symposium on Reliable Distributed Systems* 204-210. 1998.
- [41] Teich, J., H. Wallenius, J. Wallenius. Multiple-issue auction and market algorithms for the world wide web. *Decision Support System* 26 49-66. 1999.
- [42] Teich, J., H. Wallenius, J. Wallenius. A Multiple Unit Auction Algorithms: Some Theory and A Web Implementation In Electronic Markets. *International Journal of Electronic Commerce and Business Media* 9 1-7. 1999.
- [43] Teich, J. An Internet-based Procedure for Reverse Auctions Combining Aspects of Negotiations and Auctions. *Proceedings of DEXA 2000*. 2000.
- [44] Vickrey, W. Counterspeculation auctions and sealed tenders. *Journal of Finance* 16 8-37. 1961.
- [45] Vincent, A.M., J. A. Skeel. Internet reverse auctions: Valuable tool in experienced hands. *Business Horizon* **45**(4) 70-76. 2002
- [46] Vigoroso, M. Are Internet auction ready to gear up? *Purchasing* Feb11 Boston. 1999.
- [47] Wade, R.L., R. B. Harris. A Bidding Strategy. *Journal of the Construction Division ASCE* **102**(1) 197-211. 1976.
- [48] Wang, K., T.G. Eric. A study of online auction sites in Taiwan: product, auction rule, and trading type International. *Journal of Information Management* 127-142. 2002
- [49] Wang, W. L. Shill Bidding in Multi-round Online Auction. *Proceeding of the 35th Hawaii International Conference on System Sciences*. 2002.
- [50] Yue, C.Y. Decision Making Theory and Method. Science Press, Beijing. 2003.
- [51] Yong, S. K., B. J. Yum. Identification of Navigational Paths of Users Routed through Proxy Servers for Web Usage Mining. *Asia Pacific Management Review* **7**(4) 499-512. 2002.
- [52] Zhan, W. J., S. Y. Wang. Review on the Smith's mystery and development of double auctions. *Journal of Management Science in CHINA* **4**(6) 1-12. 2003.
- [53] Zhang, W. J. Auction Theory and Information Discover. *Foreign Economics and Management* **3** 99-110. 1996.

Economies of Information Consumer Commodities – An Introduction to Conceptualising Forms of Information Capitalism by Two Cases

Jussi Okkonen, Pasi Virtanen

Institute of Business Information Management, Tampere University of Technology, Finland

ABSTRACT

Change from tangible and place bound production to more flexible production of intangible information goods has challenged traditional ways of organising production. Innovative operating models have brought up also new logics of doing business, therefore one might suggest that it is not always clear which is the cause and which is the effect. The attention on general discussion is often on the nature of the product, yet at least as interesting is the meta-level, i.e. how the value is created.

The aim of the paper is to examine different operating models, i.e. ways to define operating environment. The focus is on the different models of competitiveness or how end-users are served and functionality of media service is maintained. The paper contributes to discussion on sufficient and sound ways to organise knowledge intensive value creating activities. Reflections to practice are made by examining two cases of content provisioning in digital media.

Keywords: business model, content provisioning, informationalism

1. THE CONCEPT OF INFORMATION CAPITALISM

Social and economic changes in recent decades have shifted society from industrialism to post-industrialism and now to one based on systems of information and communications. This has affected all economical entities, including people, organisations, and technologies. It has had an impact on actors at both the local and global levels. This last progression is referred as one of informationalism, where the importance of information creation and transmission is accepted as critically important to all that society represents. It does not mean that economies would no longer break down in any physical sense, nor be able to break away from physical restraints, but it does mean that these traditional limitations would be modified by growth in information and knowledge

Castells (1996) emphasises the importance of different networks for informationalism. He argued that it is essential to have a medium for seeking, exploring and researching, thus modern economic structure is dependent on information networks. He (ibid., 61) defines a new paradigm for an information-based society. In information society the information is raw material as technologies act on information not only vice versa. New technologies have pervasive effects because information is an integral part of human activity and all processes of individual or collective existence are directly shaped by the new technologies. Networking logic of any system or a set of relationships using new technologies as the morphology of the network is adapted to increasing complexity of interaction and to unpredictable patterns of development arising from the creative power of such interaction.

The role of information as a commodity and a means of production is indisputable, yet the nature of value creation process requires closer examination. Change from tangible and place bound production to more flexible production of intangible information goods has challenged traditional ways of organising production. Innovative operating models have brought up also new logics of doing business, therefore one might suggest that it is not always clear which is the cause and which is the effect. The attention on general discussion is often on the nature of the product, yet at least as interesting is the meta-level, i.e. how the value is created. In this paper the value creation is approached from traditional, say Marxist, perspective. The aim is to point out that in the production of intangible commodities the rules of tangible production do not necessarily apply. In addition to that the cases employed herein are used for the discussion of the issue of value creation in more specific.

Taking focus back to two classics of economics Smith and Marx, the emphasis was on the organisation, division of labour and ownership of capital and labour (Smith 2001, Marx 1979). The situation is still the same, the contemporary production is tied to those three relationships, but their nature is changed. The same rules or laws apply, yet there are new extensions to be considered. The main point is that for information commodities there is a very large difference between marginal cost of the first unit and the second. According to Shapiro and Varian (1999) information is expensive to produce, but very inexpensive to reproduce. That point is the first conflict between traditional economics and informationalism. Large fixed costs and low variable costs patronise centralised production and economies to scale. On the other hand technology enables production and reproduction of

inexpensive information commodities that have high surplus value to end-users.

Bischoff (2003, 33-34) points out that information technology plays significant role in value creating process in all sectors of economic and social reality. The production itself is flexible, as content (production) and distribution (reproduction) are no longer simultaneous. Moreover, technology serves as an instrument of gaining economies to scale.

Garnham (2003, 49-50) challenges the idea of increased productivity by shifting to intangible products and to using information technology. However, there is a long tradition in discussion of how the value is actually created. Notions of arbitrary value emergence (cf. value creation) can be considered false, as even voluntary participation creates both economical and social value.

By definition information capitalism refers to system where technology and intangible assets, i.e. knowledge and competencies, are used in production of information commodities. In this paper those systems can be business-to-business or business-to-consumer production systems. It is important to notice the difference between traditional systems, e.g. service provisioning and information commodity. The following section takes closer look at two different value-adding processes.

In this paper the cases are aimed to start discussion of modern information capitalism. The first case is value chain where very expensive data-set is distributed for free via Internet. The revenue logic of such activity is reverse. Consumers are offered a surplus service for free if they are willing to use the primary service, which in this case are classified advertisements in the Internet. The second case is also a value chain of public communication. In television chat the activity itself has value as entertainment, i.e. social value. The revenue logic is also similar to that with ASP (application service provider), as service provider only offers medium to customers. The customers create the content and decide how the show goes on.

2. CASE DIGITAL MAP SERVICE

ASP-model refers to business model where information system's fundamental services are fully or partly outsourced. E.g., Krajewski (2001) states that application service providers offer access to software applications and services over the Internet, virtual private networks or leased lines, and describes the phenomenon with phrases "application outsourcing", "hosted applications" and "webifying applications". In their study on software business models, Rajala et al. (2001) distinguish the following subtypes of new service providers, that the so called "Service Provisioning Shift" has recently brought into the software market: Application Services Providers (ASP),

Hosting Services Providers (HSP), Communication Services providers (usually ISP's, Internet SP's), and also content providers (e.g., CSP's) are mentioned. They form a variety of partner networks and cooperation opportunities for software vendors, and typically these companies provide hosting services for other companies in an outsourcing agreement.

Servicing and implementation model is regarded as one element of a conceptual software business model (e.g., McHugh, 1999, Rajala et. al., 2001), while the others are product development, revenue logic, sales channel and sales approach (Rajala et. al., 2001). Software implementation requires some kind of servicing, either self-service or on-site delivery. Thus it is essential to make difference between regular maintenance activities and servicing. Maintenance is a part of daily activities and based on the requirements of any system, e.g. in ASP model maintenance is mostly outsourced and customer is responsible only for the access to the Internet. Moreover, installing and configuring a system for accessing to service provided should be called servicing.

Krajewski (2001) points out three obvious benefits of the model. Firstly, savings in personnel costs, secondly, savings in other costs and, thirdly increased competitive advantage. Savings in personnel costs come mainly from the need for less support personnel as service provider maintains the system. Savings in other costs are based on the earnings logic of ASP. If service provider charges customer by transaction ASP model is always more cost-effective to some point. On the other hand, if provider charges lump sum from the customers the economies of scale ensure lower price for individual customer. Competitive advantage is perceived through up-to-date hardware and software, competent support personnel and better focus on core functions. Actually ASP is decision of make or buy, which benefits mostly small and mid-size enterprises (Raisinghani and Kwiatkowski 2001).

If ASP is examined over the life span of the system there are three main benefits. Firstly, there is no need for specific system as access to Internet provides necessary services to customers. ASP model frees also from spatial restraints, as services are available in Internet. Secondly, maintenance is outsourced, thus the risk of additional maintenance and updating are minimised. And thirdly, the system is somewhat more flexible when it is time to discard it. Thus, as it is based on contract for e.g. certain period the cost are only semi-fixed. If the hardware and software were own the costs were more fixed.

The risks of ASP are raised from the same origin as the benefits. ASP model makes customer dependent on the provider, which is an invisible relationship when everything works but if the system fails there is not much to do. Remoteness makes systems vulnerable as

even temporal breaks e.g., in Internet connection make it unavailable. If the system is provided by several ASPs the difficulties in compatibility between the components decrease usability. Especially if the supply chain is long and consists of several individually produced components, the system is sensitive to disruption. If ASP model is applied to support core functions, it is essential to maintain functionality during breaks. The difficulty of such dependability is due to that, that the ASP, also the backup system, is more or less virtual.

2.1. Oikotie – shortcut to required information

Genimap Corporation¹ is innovatively developing location-based services for the Internet and mobile information society and applies ASP for digital map services. In this paper we examine a web map service that is integrated into a media portal. The portal is called Oikotie (Finnish for 'shortcut') and it is part of Helsingin Sanomat, the biggest newspaper in Finland (www.oikotie.fi). This kind of service can be considered as an Internet extension to a printed newspaper. The Oikotie-service sells media space, i.e. space for classified advertisements to companies and private individuals. The media space is also advertisement space for companies on the site. Owned by SanomaWSOY, the publisher of Helsingin Sanomat, Oikotie contains advertisements printed in Helsingin Sanomat, and moreover advertisements in several local newspapers as well.

Oikotie contains three categories of advertisements: *apartments*, *motor vehicles* and *vacancies*. Oikotie's revenue logic is straightforward: companies and private advertisers pay for the advertisement, and the service is free of charge for the end-users looking for information. The advertisement in Oikotie may contain more detailed information compared to the newspaper ad, and include e.g. photos and links. Therefore it truly extends the advertisements on the printed newspaper. The map service is value-added service for the end-users; hence it provides usable information on the location of advertised item. For example, if the advertised item is a pre-owned car, the end-user can view a map to locate the advertiser and find a route to the destination. Moreover, if the item is a home or a job, end-user gets information about the surroundings and nearby services as well. Oikotie thus increases the information on the commodity by adding POI's² to the map and giving for example bus routes and timetables to the site.

ASP is a fairly new business model for map content providers. It offers several advantages for both the

provider and the customer. The benefits from ASP concern especially the management and cost of map service. In traditional operating model the customer was responsible of establishing a map information system, purchasing the datasets, running and administering the system and having maps updated. ASP model has several benefits e.g. economical, technical and those linked to the quality of map content. Firstly, economical benefits are gained from revenue logic of ASP. As the map service is not the customer's core service, yet it is important part of the service, Oikotie wants to focus on the business of their own i.e. selling advertisements and media space. The ASP model enables lower costs in the phase of evolving business. The reason is twofold as the capacity exceeds the use. The surplus capacity is free for other purposes for ASP and similar service can be duplicated for other customers. Secondly, technical benefits are gained from operating model of ASP. ASP model decreases maintenance costs as technology is dispersed over the value chain. Also the investments and re-investments are allocated over the network, thus need for capital, technical expertise and technological risk are dispersed more evenly. As map service provider has outsourced hosting and other support activities, it is supposed that the resource allocation is optimal. And thirdly, ASP model enables the access to most up-to-date map data, as content provider network is responsible for updating it. However, it still depends on the cycle of updates how up-to-date the maps are.

The value chain of Oikotie consists of three layers, which we call the component provider layer, the ASP layer and the front-end layer. The component provider layer contains software providers and content providers. As the map service provider integrates both the content from different sources into one service, and the information system using software components from different vendors.

The ASP-layer is a middle-layer that enables the use of map service to customers. ASP-layer contains map service by Genimap, hosting service, i.e. the physical place for database server with 24-7³ secured service, and communication service, i.e. channel to access data. Genimap concentrates on the development and maintenance of map services, other services on the ASP-layer are outsourced. The customer's portals are placed on the front-end layer that is accessible for the advertisers and end-users. It consists of different portals, which are connected by some common content. For example, the Finnish real estate agent Huoneistokeskus advertises in Oikotie, but most of the content is in its own portal. All of the front-end providers are not linked, but they all utilize map service provided by Genimap.

¹ See Okkonen, J. & Salo-Merta, L. 2002 for more detailed discussion on ASP model in digital map services

² POI=point of interest

³ 24-7=24 hours a day, seven days a week

3. APPLICATIONS UTILISING MOBILE COMMUNICATIONS – CASE TV-CHAT IN FINLAND

By now there are numerous entertainment applications that are used via mobile communication devices. Mobile communication devices can be interpreted in most cases as mobile telephones, but also communicators. On market there are games to be played in mobile phones, games to be downloaded from various web sites to the mobile phones and even games that can be played by using mobile phones as interfaces and television screen and program as display. One perhaps more communicative new way of doing business is to use television and mobile phone in interaction. In this sector there are at least two types of applications in use in Finland. Firstly there is the TV-chat, which is introduced more closely in this paper and secondly the game sector.

By now at least every Finn has probably seen these short messages that are seen on TV-screens, called chats or chat boxes. The phrase is also adapted to Finnish everyday language “chatti” from the English *to chat*. The action itself is quite straightforward. One sends a SMS-message⁴ to a channel specified number and the sent messages are then aired in almost real time. The phrase “almost real time” is used, because the difference between the more quiet moments (=less TV-viewers and thus less *chatters*) and the busy rush hour can be several minutes.

As the product was new, one might have wondered what it was all about. After a short, closer look at things, it may be stated that a new entertainment for the youth and a lucrative moneymaker for the TV-channels was born. Lucrativeness is constituted by the fact that the content of this service is provided by the users themselves, the service provider, TV-channel, only offers the possibility to do this against a fee and airs the messages.

TV-chat is real time communication between two or more persons through mobile phones, computers and the Internet. *The chat area* is like a virtual town square to which people gather together to discuss with each other. The communication is primarily made by typing notes to one another with and through the mentioned tools (mobile phones, computers). To enable *the chat* also a server is needed. The chat service is located at the server on the service providers' premises and this server then upholds the *chat area*. (Vuorinen & Hendrell 2003)

There is a control or a controller on behalf of the service provider to monitor the SMS-messages and that the messages follow the given, sometimes unwritten

rules. The rules include among others following: no contact information is to be shown, no abusive or offensive messages are aired whether it is about gender or race. The TV-companies are bound by the Finnish law and general ethics. The messages may be edited, shortened or not aired at all, depending how the controller sees fit. Mainly the messages are displayed as they were written so they are not corrected concerning their writing.

The chat services differ from one another to some extent. At some TV-channels the controller takes part in chat conversation and comments on the messages. Also some channels display live footage taken by a web cam of the controller simultaneously on the rim of the screen. It seems that the controllers have become sort of celebrities themselves and that some of them are more popular than others.

TV-chat started in the Internet and its model is a Finnish originated IRC (Internet Relay Chat), which enables web-users' multiple conversations on different channels (Note the difference in Finnish terminology to chat and to irc, *chätätä* and *irkata*). IRC was developed in Finland in the summer of 1988 by Jarkko Oikarinen of the University of Oulu.

Chat found its way to the television through the text-television. For example during the fall 2002 Finnish TV-broadcaster MTV3 had five text-TV pages through which the messages could be sent 24 hours a day. All of these pages are controlled like the TV-chats (Kivimäki & Saarinen 2001). The main difference between a text-TV-chat and a TV-chat is that one can send contact information through the text-TV-chat and that the control does not comment on the messages. Of these two the TV-chat is more popular in Finland (Kangaspunta & Huusko 2002). Sometimes the two are used in connection with each other.

First to start with this hype-to-be was a TV-channel called TVTV! (which changed its name in autumn 2001 to SubTV) in June 2000 (Kuorilehto 2003). Now the chat-service is provided by major commercial TV-channels such as MTV3, TV4 and also TV Tampere and SportChannel. At the beginning the service was just meant to fill the odd hours of the night when there was no program on. As the whole concept grew more popular, the services spread to other hours of the day. Now they are offered also in the mornings and in the afternoons, summing up to multiple tens of hours each week. First night the MTV3 had their TV-chat service, they received thousands of messages almost blocking the service. In the morning when the service was about to close there was a queue of hundreds of messages still to be displayed (Ryöti 2003). It was mainly the youth who sent each other greetings, “philosophical” notes and quotations and seemingly irrational comments such as “Tappara rulez” (Finnish ice-hockey team). Today the services display a “no-more-messages” notice,

⁴ SMS=short message service in mobile telephoning context

when their air-time threatens to end before all the messages can be displayed (Ryöti 2003).

The sending of messages is possible at least for the subscribers of teleoperators DNA, Radiolinja, Sonera and Telia. Price of a message to a chat service varies a little. It generally is 0,84 euros (MTV3 chat, 28.10.2003) another source declares that for Radiolinja subscribers the prices are 0,59 euros for the SMS-messages (the "normal" text message) and 1,00 euros for the MMS-messages⁵ (TTV 02.11.2003).

3.1. The technical side of the chat services

The chat begins with a normal SMS-message from one's own phone to a given number. These chat-numbers are TV-channel specific and they differ from the "ordinary" telephone numbers significantly (for example 173222 instead of 040-5556789), so there is little or no risk of accidentally calling these numbers. The messages are directed from one's "own" tele-operator to central for the SMS-messages and further to a third party, another service provider, which claims to be a value-added service for communication networks. One such service is INA Finland Oy in Tampere (Hynnä 2003).

A server hosted by the third party service provider then 'chops up' the message and separates from it the information not needed or wanted for the chat purpose. Contact information of the sender is this kind of information among other things. After this chopping up the message is then sent to the TV-chat service provider, to TV-channel, there it ends up at the computer of the chat controller. In this stage the chat controller can edit, comment or remove the message as shown earlier. Finally the chat controller approves the message, which is the most common case. (Hynnä 2003)

The approval sends the message back as "raw text" to the third party service provider again. There it goes onto another server, where the message is converted to a form in which it can be sent on air. The conversion is actually already done, in this stage a previously created alias can be added to the message or a specific colour or some other chosen options and parameters. After these procedures the message then returns for the final time to the TV-channel to a so called "air-server", which actually is a web-based screen that only sends the message on air. At this final stage no alterations are possible. (Hynnä 2003)

Previous systems were not so, at least seemingly, complicated. However, practice has shown the supremacy of these newer systems. They are better and easier to maintain and uphold, because each machine

does only one part of the whole. The system is easier altered and programmed, for example creation of aliases. Also in case of problems the damages can be bound not to whole system but to a more controllable part of it. (Hynnä 2003)

The provider of the actual chat service must have adequate communication channels. However this presents a lesser problem because the text messages present a rather small burden to the net connection. In the interviewed company the connection was a 2 Mb's (megabit's, millions of bits per second) connection to prevent any possible collisions, which is to make sure. This was said to be far more than adequate. They also had an automated service which checks the new messages twice a second. This stage of automatization and these connections result in for a sent message to be shown on the chat controller's screen in five seconds from sending it and in another five to be aired. These numbers depend somewhat on the fact how quick the chat controller is to accept the message, how busy the networks are and how many messages are sent simultaneously. (Hynnä 2003)

Finally, one interesting fact related to this activity is that according to the Finnish law, all messages must be kept for a certain period of time. As time goes by the storage capacity of the chat service provider must also be considerable. This is yet another question for the TV-chat provider to answer, but this area is marked out in this paper.

4. CONCLUSIONS

Comparing the two cases some contradictory notions can be pointed out. The most important notion on digital business is, that content is often costly to create but inexpensive to reproduce. That is the fact along the maps, but untrue with the TV-chat (cf. Shapiro and Varian 1999, 21-22). In fact there is a brilliant revenue logic on idea that customers pay for the joy of providing media content. It seems that in both cases the revenue logic is bound around information content, but the relationship between the "true" content and surplus service is blurred. In the map-case the once produced material is sold over and over again whereas in the chat-case the content is actually not sold at all.

In the case of TV-chat the idea is that people entertain themselves by public communication. The platform that enables such activity is domain or service provider not the content. Such operating model is the most inexpensive one, yet it is still productive. Taking TV-chat as an example of production, there is fuzziness about the mutual relationships in the value chain. The beginning of the value chain is clear. There is hardware and software that the systems consist of, but when the content is added the fuzziness appears. The show, the TV-chat in itself, is co-operation between viewers/chatters and the controller. It might be claimed

⁵ MMS=multimedia message service in mobile telephoning context

that people are willing to pay relatively high fee for participation. The actual media content is free for all those who possess a television set.

TV-chat can be considered a sort of exception that confirms the rules, but still it raises interesting questions on relationships in production. For the chat service there is logical explanation on people's willingness to have their time on TV. As the mobile phone penetration in Finland is close to 90 % (Digitoday 2003) and there are some 3,6 million TV-receivers in the country there is bound to be people willing to take part in this. However the publicity in that forum is still very limited. Other notion is that people do not pay for viewing the chat messages, but posting them.

In the case of digital maps the explanation for such model is that as the data-sets already exist and there is a lucrative way to use them for minimum marginal cost, the operating model is ready. Moreover, if digital maps were only content in the service there might be fee for using them. There are several other examples of such operating models where people get something free of charge, but according to famous tag by Milton Friedman "there is no such thing as a free lunch".

These two cases point out that if there is possibility to produce and reproduce information with low costs, the information can be used as an independent service or a surplus service. The revenue does not necessarily come straight from the service, more likely from other sources. Information can, and will, be used in different combinations, thus the information goods are often complex combinations and layers of information and knowledge. What seems to be clear at the first sight, can be very complicated at the second.

REFERENCES

- [1] Bischoff, J. (2003) Moderni kapitalismi ja informaatioteknologia pp. 31-39 in Heiskanen, J and Mäntylä, J. (eds.) (2003) MarxIT – Informaation kapitalismin kriittistä tarkastelua. Karl Marx-seura ry, Helsinki
- [2] Castells, M. (1996) The Infomation Age. Economy, Society and Culture. Volume I: The Rise of The Network Society. Oxford, Blackwell
- [3] Digitoday, 2003. Yhä useampi vain matkapuhelimen varassa.
- [4] http://www.digitoday.fi/showPage.php?page_id=12&news_id=3115. Read 10.08.2004
- [5] Garnham, N. (2003) Tietoyhteiskunta: Teoriaa vai ideologiaa? Kriittinen näkökulma teknologiaan, koulutukseen ja työllisyyteen informaation aikakaudella pp. 41-57 in Heiskanen, J. and Mäntylä, J. (eds.) (2003) MarxIT – Informaation kapitalismin kriittistä tarkastelua. Karl Marx-seura ry, Helsinki
- [6] Heiskanen, J. and Mäntylä, J. (eds.) (2003) MarxIT – Informaation kapitalismin kriittistä tarkastelua. Karl Marx-seura ry, Helsinki
- [7] Hynnä J., Director TV Tampere. Interview 03.11.2003.
- [8] IRC.org. www.irc.org/history_docs/jarkko.html, 05.11.2003
- [9] Kangaspunta, S., Huusko, M. 2002. Yksisuuntaista odottelua, Terveiden ja hyvinvoinnin yhteisötelevisio –tutkimusprojektin väliraportti II Tammikuu 2002. <http://www.uta.fi/jourttutkimus/20402raportti2.pdf>, 29.10.2003
- [10] Kivimäki A., Saarinen J. 2001. Televisiochatit: Äärimmäistä ajanhukkaa vai alku jollekin suurelle? Tiedotustutkimus 3/2001.
- [11] Krajewski, K. 2001. What's Up with ASPs? Association Management, May 2001
- [12] Kuorilehto, M. 2003. Chat-kielen piirteistä. Kielikuvia 1/2003. <http://www.uta.fi/jarjestot/nsseura/kielikuvia/kielikuvia-1-03.pdf>, 29.10.2003
- [13] Marx, K. (1979) Pääoma, Osa 1: Pääoman tuotantoprosessi. Kustannusliike Progress, Neuvostoliitto
- [14] McHugh, P. 1999. Making it big in software. Professional Book Suppliers. Oxford, England
- [15] Okkonen, J. & Salo-Merta, L. (2002) Digital Maps Go ASP - Application Service Provider Model Applied to Digital Map Services. Europrix 2002 Tampere 14.-16.2002
- [16] Raisinghani, M. and Kwiatkowski, M. 2001. The Future of Application Service Providers. Information Strategy, summer 2001.
- [17] Rajala, R., Rossi, M., Tuunainen, V. K., Korri, S. 2001. Software Business Models, A Framework for Analyzing Software Industry. Technology Review 108/2001. TEKES.
- [18] Ryöti M., Planning Manager, MTV3. Interview 06.11.2003.
- [19] Shapiro, C and Varian H. R. (1999) Information Rules. A strategic guide to the network economy. Harvard Business School Press, Boston
- [20] mith, A. (2001) Of the Division of Labour. pp. 37-41 in Shafritz, J. M. and Ott, J. S. (eds.) (2001) Classics of Organization Theory, 5th edition. Wadsworth Group, Belmont.
- [21] UKES, 2002. Televisiot tulipalojen aiheuttajina. TUKES turvatekniikan keskus. http://www.tukes.fi/sahko_ja_hissit/esitteet_ja_opaat/TV_palo_suomi_5_02.pdf. Read 10.08.2004
- [22] Vuorinen, V., Hendrell, H. 2003. Chatmaailma.com, chat-tutkimus. <http://www.chatmaailma.com/chat/abc/perustiedot.php>, 29.10.2003.

Effective and Strategic Use of Electronic Reverse Auctions

Chia Yao Lee¹, Brian Corbitt²

¹ School of Information Systems, Deakin University, VIC 3125, Australia.
chia.lee@deakin.edu.au

² School Graduate School of Management, Shinawatra University, Bangkok, Thailand, and
School of Information Systems, Deakin University, VIC 3125, Australia.
bc250604@yahoo.com.au

ABSTRACT

This paper reports the findings of an empirical study of Electronic Reverse Auctions (e-RAs). This study seeks to better understand the underlying and emerging issues in e-RAs to comprehend the implications of e-RAs on organizational procurement. The study contributes to practice and research by providing insight into emerging e-RA adoption issues, and by discussing strategies for improving e-RA processes, thus improving the effectiveness of e-RAs.

Keywords: Electronic Reverse Auctions, Electronic Procurement, Electronic Sourcing

1. INTRODUCTION

In recent years, the use of Electronic Reverse Auctions (e-RAs) for sourcing and procurement has attracted significant interest from the public and private sectors [1, 2, 3, 4, 5]. In 2001-2002, up to \$40 billion worth of procurement transactions were conducted via e-RAs [6]. Beall et al. reported that e-RAs generate an average savings of 15% on the prices paid for goods, and decrease procurement cycle time by up to 90% [7]. GlaxoSmithKline, a major pharmaceutical supplier, achieved cost savings of \$165 million, from a series of 190 e-RA events worth \$912 million [7]. Procurement represents a major spend area for organizations. Modest savings of a few percentage points can translate into multi-million dollar savings in the longer term. e-RAs promise to improve the bottomline of organizations by reducing the prices paid for supplies, as well as lowering the cost of transaction through process improvement [7].

It is not surprising that purchasing executives often face questions such as “Why should their organization adopt e-RAs for sourcing supplies?”, “Which supplies should they source through e-RAs?”, “How could organizations leverage the benefits of e-RAs?”, and “What are the potential challenges associated with e-RAs?” These questions do not have conclusive and generalizable answers as the outcome of e-RAs depends on factors such as the amount of planning and preparation, the existing relationships between the buyer and seller, and the attributes of products that are to be purchased [8]. Opponents of e-RAs are doubtful of the sustainability of e-RA benefits. They argue that incumbent suppliers view e-RAs as divisive; a mechanism that damages well-established relationships between buyers and sellers [9]. Supporters and critics of e-RAs argue passionately for and against e-RAs, clouding the discussion over how to leverage the strengths of e-RAs, and to avoid the potential pitfalls. This study aims to identify, and better understand the emerging issues in

e-RAs. The study addresses the research question:

“How should organizations adopt e-RAs effectively?”

This paper describes a case study of a series of six e-RA events hosted by a major multinational organization in Australia. In answering the principal research question, the study is also motivated by additional guiding questions that ask: *What’s new? So what? and Why so?*

2. BACKGROUND OF ELECTRONIC REVERSE AUCTIONS

Kambil and van Heck describe e-RAs as an auction process in which a single buyer places a request for a quote for a product it intends to purchase, and multiple sellers (the suppliers) place bids to sell to the buyer [10]. The bids are placed until the lowest price is discovered when the auction closes at a pre-determined time. e-RAs are also commonly known as “Online Reverse Auctions”, “Downward Price Auctions”, and “Electronic Procurement Auctions” [11, 12]. The e-RAs idea was conceptualized by Glen Meakem, who subsequently founded Freemarkets Inc [2]. e-RAs differed from English Forward auctions in that the roles of the buyer and seller are swapped, and a decremental bidding mechanism is used. They are used in procurement mainly for evaluating the offerings of competing suppliers, and for determining the purchase price of goods purchased. According to Kambil and van Heck’s Basic Trade Processes framework [13], e-RAs fulfill the business process phase known as “valuation” which is described as

“negotiating and discovering a purchase or sale price for a product. A variety of different price discovery and bidding processes exist that differentially attributes costs to buyers, sellers and intermediaries. ... Price discovery mechanisms can also be biased to shift surplus from the trade to specific transaction stakeholders.”

Literature supporting the e-RAs model typically relies on empirical evidence that illustrates the quantifiable financial savings that can be attributed to e-RAs. Jap asserts that organizations adopt e-RAs to achieve three goals – (i) immediate financial savings, (ii) process efficiency gains, and (iii) enabling capabilities such as global and multiple sourcing [5]. In a similar vein, Stockdale and Standing found that apart from direct financial savings, e-RAs also help organizations standardize procurement processes, and develop staff purchasing skills [14]. Kambil and van Heck concluded that apart from achieving financial savings, e-RAs enable buyers to gain useful market information, such as suppliers' bottomline or reserve price [10].

Critics and sceptics of e-RAs argue that the e-RA model damages buyer-supplier relationships. To win contracts through e-RAs, suppliers are pressured to sacrifice profit margin to levels that are unsustainable in the longer term. It was found that e-RAs achieve local optimization of business processes, an effect that leads to chronic underutilization of internal and external resources [11]. In a subsequent study, Emiliani and Stec also found that incumbent suppliers attempted to charge buyers higher prices to compensate for their participation in e-RAs [15]. Another strong argument against e-RAs relate to the importance of price in buyer-seller relationships. Some attributes of a product, such as delivery, settlement, after-sales support, do not feature prominently in e-RAs negotiations.

These apparently different perspectives may suggest that the e-RAs model was not fully understood by organizations, practitioners and researchers. Also, e-RAs continue to undergo evolution, with new challenges emerging every so often. This study addresses some of the above issues. Without placing too much emphasis on literature and testing of existing theories, this study concentrates on identifying the emerging issues in e-RAs from the perspectives of buyers and sellers. The study recommends strategies that target the source of the problem encountered in e-RAs [16].

3. METHODOLOGY

The study describes an exploratory study that draws data collected from a series of six e-RA events. An interpretivist approach was adopted, as many of the issues emerging may be incomplete and continue to evolve. The interpretivist approach allows for a belief that the reality is subjective, and is the product of social construction, to be viewed upon by humans, who are the social actors [17]. In capturing the properties of the subject, reality is interpreted according to the different beliefs, value systems, and consciousness of the social actors.

The study does not intend to develop a metric for determining the descriptive variables, nor is it our aim

to use inferential statistics for producing generalizable and reproducible outcomes. The aim is to develop a deeper level of understanding of the e-RA phenomenon, and aid the understanding of important concepts, themes, and their implications. Data was primarily collected through informal, semi-structured interviews. The interview questions were developed from an understanding of the constituent nature of the relationships between the stakeholders and the phenomenon, which are garnered from existing literature. Interviews were conducted with executives of the focal organization and its suppliers. Publicly available literature was used as a secondary data source. Data was then synthesized using a hermeneutics cycle to determine emerging concepts, and also for summarizing common themes.

The focal organization in the case study represents an Australian branch of a multinational organization. The parent business has a diversified portfolio of businesses, but the focal organization operates primarily in the financial sector. Its financial products are either resold by other businesses in Australia, or sold directly to the general public. Confidentiality restrictions prevent details that may lead to the identification of the participants from being included in this paper. For the purpose of the study, the focal organization is referred to as FinanceCo. FinanceCo conducted a series of e-RAs in 2000 and 2001, of which six are included in this study. e-RAs were not intended to complement FinanceCo's existing procurement channels. This paper reports the observations made, and suggests possible explanations for describing the phenomena.

4. RESEARCH FINDINGS

Analysis of the interview data and synthesis of secondary data yields the following emerging themes. The themes are presented here in no particular order of importance.

4.1 The Negotiation Parameters and Lotting Strategies

The adoption of e-RAs for procurement limits the negotiation parameters available to buyers and sellers. The only negotiation parameter that suppliers could vary in e-RAs is the purchase price. Traditionally, other non-price parameters like settlement, logistics and inventory management could be negotiated upon. In preparing the tender for the e-RAs, FinanceCo conducted substantial research on the product and the potential suppliers. It needs to find out the product's specifications, costs, the market demand and supply, and the attributes of the potential suppliers. This is to ensure that the e-RA tender document covers all possible aspects of the product. The performance, quality, and expectation of the product are usually benchmarked to industry-backed standards, or mutually agreed upon standards. Once finalized, these negotiation parameters

become non-negotiable to ensure that in evaluating and comparing the bids, a fair and just environment exists for “comparing apples with apples”.

In the e-RA event for courier and forwarding services, it was observed that a particular supplier attempted to differentiate their offerings. The supplier was willing to match an existing lowest bid by providing services that are higher in quality, but was not keen to lower the purchase price. Superficially, this bid appear to be of better value than competing bids, but FinanceCo adhered to the terms and conditions for the e-RAs strictly, and advised all bidders that all bids that satisfy the required specifications are being treated equally. In doing so, FinanceCo signaled that suppliers could add value to a bid by offering over-specifying a product, but will not win the auction if a lowest price bid was not placed.

The transparency of information in e-RAs raised other challenges associated with the absence of non-price negotiation parameters. e-RAs led to the bundling and unbundling of purchases. Components of a purchase which were bundled in the past could be purchased individually due to the increased ease in sourcing, and the ability to individually determine the price of the components. Commoditization of products also occurred, with the emergence of no-frills products, with suppliers removing free of charge value-added services in order to lower prices. For example, in the auction for hotel accommodation for staff, the inclusion of breakfast was halted to simplify the comparison of standard room accommodation. The unbundling of the purchase enabled FinanceCo to view the cost structure of its suppliers, as well as optioning-in (or -out) non-essential features of a product. Nevertheless, a re-bundling of value-added services was also observed. In the e-RA for computing peripherals, FinanceCo added a new service – a requirement for suppliers to dispose of superceded old peripherals. In e-RAs, the buyer has to be explicit in specifying delivery, settlement, product specifications and other special requirements when preparing the tender document. It is not surprising that the competitive nature of e-RAs encourages suppliers to do away with “frills” in their products, or charge for such “frills” separately.

In the absence of non-price negotiation parameters, FinanceCo has used alternative approaches for purchasing a “basket” of products. Instead of purchasing a large lot of identical products, FinanceCo was able to purchase products of different quality or performance separately. The breaking up of large purchase lots into several smaller ones meant that the smaller lots could either be sourced from the same supplier, or from multiple suppliers. This represented an opportunity that was not exploited before the implementation of e-RAs as the search costs involved in sourcing various components individually often outweigh the cost of purchasing a bundled product.

In terms of aggregating purchase volume, e-RAs enabled FinanceCo to purchase in larger lots at some occasions by combining its purchase with other arms of its parent business. For example, in the e-RA for hotel accommodation, FinanceCo had aggregated its purchases with other business divisions of its parent company. By combining their purchases, they took advantage of economies of scale and achieving critical mass for less important purchases. From a supplier’s perspective, selling to FinanceCo via e-RAs represents an opportunity for the supplier to supply different components of a larger product, and where possible, supply a fully-packaged product.

The observation regarding the change of behaviour in FinanceCo’s purchasing indicate the possibility in optimizing its procurement. Apart from increasing the competition amongst suppliers that could lead to lower purchase prices, FinanceCo gained a better understanding of its purchases, enabling it to optimize purchasing decisions to take advantage of individual supplier’s capabilities. The common rule of thumb for organizations sourcing through e-RAs is to purchase from the lowest-priced suppliers. However this poses a new challenge for FinanceCo in the determination of “value for money” of purchases. The absence of non-price negotiation parameters could make suppliers market their products opportunistically. For example, products that require after-sales support may be marketed with a low initial purchase price, but the cost of subsequent after-sales support would be priced higher to counter the initial low purchase price. While such practices are despised, the buyer has to understand “you get what you paid for” and e-RAs are principally a more effective mechanism for determining prices, and possibly the allocation of resources. Buyers will also need to take into account the Total Cost of Ownership of a product as the initial purchase price could only be the tip of the iceberg. This issue creates complications for the sourcing of indirect supplies, whereby repair, maintenance, upgrading and disposal of equipment incur significant costs. For commodity-like direct supplies, the Total Cost of Ownership could be less of a worry, but the quality, delivery and settlement flexibility should be taken into account in the purchase price. Additionally, the value of collaboration between suppliers and the buyer in product research and design could be overlooked in e-RA transactions. Flexibility in delivery, inventory management and settlement were important reasons why organizations adopted Supply Chain Management, and organizations sourcing through e-RAs need to consider the implications from the loss of such collaboration. Additionally, the intangible attributes of the product, unique capabilities of the suppliers and process improvement (e.g. Just In Time manufacturing) cannot be expressed in terms of prices, nor can they be assumed to be the same. Thus, e-RAs present a significant challenge for buyers and sellers in pricing intangible attributes of products, and the attributes of the suppliers.

4.2 The Timing of Electronic Reverse Auctions

In determining whether to source supplies through e-RAs, FinanceCo first evaluated various attributes of the product involved. A major consideration was given to the number of suppliers for the product, and their tendency to bid for the contracts competitively. If a product is only sold by one or two suppliers, FinanceCo would decide against sourcing it through e-RAs. When there is a small pool of suppliers, the suppliers tend to know each others' operations and cost structure well, and will only be willing to match their competitors' bids. Under such circumstances, the e-RAs business model will not deliver substantial benefits to the buyer in terms of price reduction. Similarly, if a product is sold by many suppliers, but none has spare capacity to satisfy the volume required by FinanceCo, then e-RAs will not generate significant competition.

A strategy that FinanceCo has applied to raise the interest of suppliers to place competitive bids and thereby increase competition in the market is through aggregating the purchases of all its branches, and at times, its sister companies. At the same time, FinanceCo evaluated the market for a particular product, determining periods of excess capacity. After analysing the purchasing patterns of other organizations, FinanceCo scheduled its e-RA events before other organizations purchase, ensuring that potential suppliers have not made substantial commitment to other customers. By coordinating its purchasing timing to pre-empt other organizations, FinanceCo was able to get the best participation rate and possibly the best outcome from its e-RA events. Also, suppliers who wish to secure future business when their existing contracts terminate would be able to participate in e-RAs without being caught in a capacity shortfall. This two-pronged approach in enlarging a purchase order, and choosing a period of excess supplier capacity enabled FinanceCo to maximize competition among potential suppliers.

However, FinanceCo was aware that e-RAs should not disadvantage suppliers to the extent that healthy competition is diminished, especially during periods of economic downturn. Reliable supplies require profitable suppliers in the long term. Thus, FinanceCo treads a fine line in increasing the competitiveness of its suppliers, while not driving the suppliers to oblivion. FinanceCo also constantly searches for new suppliers that have the expertise and capacity to satisfy its needs.

4.3 Education and Awareness

e-RAs introduce the notion of the shift from place to space. It enables suppliers to bid for the tenders remotely through the Internet instead of physically attending RA events. While such practices have been common in seller-hosted forward auctions, it is not often that suppliers have to bid against their competitors in a live environment to win the privilege of selling their

products. To prevent suppliers from colluding, yet to ensure that suppliers do not view e-RAs simply as a mechanism to obtain price quotations, the host of e-RAs must adhere to mutually agreed terms and conditions, and provide some level of transparency. As observed in all of the six e-RA events, suppliers' pricing strategies were reactionary. They were not willing to be the price-leader. Rather, suppliers were keen only to match the prices of the price leader. However, the way that e-RAs were designed ensures that they deliver most benefits to the lowest price bidder. To further enhance incentives for suppliers to participate in e-RAs, the e-RA sponsors (either the buyer, or the buyer's representative) have to educate potential suppliers. This is to increase awareness of participating bidders about their obligations and rights, as well as about acceptable practices by the buyer and sellers.

To ensure that suppliers do not experience technical difficulties or have inadequate infrastructure for participating in the e-RAs, FinanceCo organized a trial run for the auction on the day before the actual auction. However, to distinguish the trial run from the real auction, FinanceCo reminded the bidders not to place bids based on their true pricing strategies. The purpose of the trial auction was to provide bidders with a "simulated" experience, and to iron out potential misunderstandings of the e-RA process. FinanceCo also ensured suppliers that it did not embark on e-RAs purely for the purpose of price discovery, but that it was serious about switching suppliers.

The training and education of bidders raise their awareness of the binding nature of e-RAs. They learn that although e-RAs facilitate price determination, bids placed are not retractable. Therefore bidders will need to have a clear knowledge of their own cost structure before the e-RA to ensure that the bids they place are achievable and realistic. In addition, the training and education exercises also reduce the anxiety and emotional experiences bidders may face in e-RAs. Bidders are encouraged to place bids rationally, and not be affected by a spur of the moment emotion, or be intimidated by competitors' bids. The training and education efforts also ensure a consistent familiarity of participants with the auction process.

The above are observations that were picked up over several e-RA events. The importance of some of these issues may change depending on the e-RA event. Nevertheless, the host /sponsor of e-RAs has to consider all possible eventualities, and come up with practical and equitable solutions.

5. IMPLICATIONS FOR RESEARCH AND PRACTICE

The present study highlighted several key issues that are emerging in e-RAs. A major challenge for buyers in deciding whether to implement e-RAs in sourcing and

procurement would be the ability to pre-determine the negotiating parameters, enabling price to be the sole negotiation parameter. But often, this is not the case; if the buyer continues with e-RAs, then a set of different pricing and costing methodology is required. The pricing and costing methodology will have to take into account the Total Cost of Ownership, as well as the ability to separate complex pricing strategies into their individual components. While this may appear as an attempt to commoditize non-commodity supplies, organizations that pursue the outsourcing approach are facing similar challenges to individually price components of a complex bundle of products. Multiple-attribute e-RAs [18] that take into account the various attributes of a product auctioneered could overcome some of the challenges of a price-based e-RA. The different attributes are assigned different weightings, and a scoring system that calculates the total "worth" or "value" of a product could substitute the use of price as the negotiating parameter, i.e. develop a negotiating parameter that represents the attributes of a product better than price.

The second issue is the ability for e-RAs to produce "Win-Win-Win" situations, that is, to reward more than just the buyer and the winning supplier. Efforts to reward non-winning bidders will encourage their future participation. Non-winning bidders should be made to appreciate that by participating in e-RAs, they gain access to market price and competitors' information. A possible solution to this problem is a multiple-sourcing e-RA model. Instead of awarding the whole auction lot to the winning supplier, the buyer may find that a 70/30 or 60/40 split between the winning bidder and the second lowest bidder could provide additional motivation for suppliers to participate in e-RAs, as well as encouraging suppliers to bid competitively. Attempts to source supplies through variable lotting strategies indicate efforts by the buyer to spread its purchase among several suppliers.

Finally, organizations need to recognize the role of e-RAs in procurement and sourcing in price-determination and supplier selection. Supplier selection is a process that is not totally influenced by a product's price. In e-RAs, the pre-auction qualification process represents initial supplier selection whereby the unreliable suppliers are weeded out, and capabilities of suppliers verified. Organizations also need to recognize the variance among the different types of procurement they undertake. Direct and indirect supplies require different procurement strategies. In preparing for an e-RA, the buyer should consider whether it would be willing to place competitive bids if it were placed in the position of a supplier.

6. CONCLUSION

This paper summarizes key observations from a case study of e-RAs. Using empirical data collected from

stakeholders involved in e-RAs, this paper provides insight into the emerging issues in e-RAs and suggests possible explanations for describing the observations made. While not attempting to generalize the observations, the explanations provide cues for developing strategies for overcoming specific challenges that arise from e-RA implementation.

The study's findings contribute to practice particularly in the determination of e-RA negotiation parameters. Current practice to use price as the sole negotiation parameter introduces substantial limitations for e-RAs.

A multi-attribute weighted scoring system could provide a more complete and definitive indicator of a product's attributes and properties, and hence could be the basis of a single and more holistic negotiation parameter.

A major limitation of the present study is that the e-RA events involved the sourcing of indirect supplies. Direct supplies are more critical to an organization's operations, and the applicability and effectiveness of e-RAs may be limited as incumbent suppliers could have more control over the buyer-supplier relationship, preventing the buyer from switching suppliers without incurring significant costs and risks.

REFERENCES

- [1] Emiliani, E. M., "Business-to-Business online auctions: Key issues for purchasing process improvement", *Supply Chain Management*, Vol. 5, No. 4, pp176-186, 2000.
- [2] Tully, S., "The B2B tool that really IS changing the world", *Fortune*, Vol. 141, pp312-145, 2000.
- [3] Desai-Sarnowski, A., and Murzyn, P., "Which electronic solicitation to use? Reverse-Auction or Sealed Bid?", B2emarkets, URL: www.b2emarkets.com, 2002.
- [4] Turley, S. L., "Wielding the virtual gavel - DOD moves forward with Reverse Auctions", *Military Law Review*, Vol. 173, September, pp9-67, 2002.
- [5] Jap, S., "An exploratory study of the introduction of Online Reverse Auctions", *Journal of Marketing* Vol. 67, No. 3, 2003.
- [6] Jap, S., "Online reverse auctions: Issues, themes, and prospects for the future", *Academy of Marketing Science*, Vol. 30, No. 4, pp506-525, 2002.
- [7] Beall, S., Carter, C., Carter, P. L., Germer, T., Hendrick, T., Jap, S., Kaufmann, L., Maciejewski, D., Monczka, R., and Petersen, K., *The Role of Reverse Auctions in Strategic Sourcing*, Tempe, Arizona, CAPS Research, pp1-87, 2003.
- [8] Lee, C. Y., Corbitt, B. "A Stakeholder-Benefit perspective of reverse auctions", *Proceedings of the 7th Americas Conference on Information Systems (AMCIS)*, Boston, 2001.
- [9] Emiliani, E. M., and Stec, D. J., "Aerospace Parts Suppliers' Reaction to Online Reverse Auctions", *Supply Chain Management*, Vol. 9, No. 2, pp139-153, 2004.
- [10] Kambil, A., and van Heck, E., *Making markets: How firms can design and profit from online auctions*

and exchanges, Boston, Harvard Business School Press 2002.

[11] Emiliani, E. M., and Stec, D. J. "Realizing savings from online reverse auctions", *Supply Chain Management*, Vol.7, No. 1, pp12-23, 2002.

[12] Kaufmann, L., and Carter, C. R., "Deciding on the mode of negotiation: To auction or not to auction electronically", *Journal of Supply Chain Management*, Vol. 40, No. 2, pp15-26, 2004.

[13] Kambil, A., and van Heck, E., "Reengineering the Dutch flower auctions: A framework for analyzing exchange organizations", *Information Systems Research*, Vol. 9, No. 1, pp1-19, 1998.

[14] Stockdale, R., and Standing, C. "The effect of Business to Business online reverse auctions on buyer-supplier relationships", *Proceedings of the 14th*

Australasian Conference on Information Systems (ACIS), Perth, Australia, 2003.

[15] Emiliani, E. M., and Stec, D. J., "The inevitability of conflict between buyers and sellers", *Supply Chain Management*, Vol. 8, No. 2, pp107-115, 2003.

[16] Sawhney, M., "Forward thinking about reverse auctions", *CIO*, URL: www.cio.com.au/pp.php?id=711454420&fp=4&fpid=4, 14 July 2003.

[17] Klein, H. K., and Myers, M. D., "A set of principles for conducting and evaluating interpretive field studies in information systems", *MIS Quarterly*, Vol. 23, No. 1, pp67-94, 1999.

[18] Talluri, S., and Ragatz, G. L., "Multi-attribute reverse auctions in B2B exchanges: A framework for design and implementation", *Journal of Supply Chain Management*, Vol. 40, No. 1, pp52-60, 2004.

How E-Ready is E-Sri Lanka?

Velauthapillai Dhayalan¹, Reggie Davidrajuh²

¹ Computer Science, Bergen Univ. College, Norway, vdh@hib.no

² Electrical & Computer Eng., Stavanger Univ. College, Norway, Reggie.Davidrajuh@tn.his.no

ABSTRACTS

First, this paper assesses e-readiness of Sri Lanka using a measuring tool that utilizes 52 socio-economic indicators. Second, based on the assessment, this paper suggests methods for improving e-readiness of Sri Lanka. The uniqueness of this paper lies in the quantitative analysis of the e-readiness of Sri Lanka, whereas many other works engage in qualitative exposure.

Keywords: E-government, e-Sri Lanka, e-readiness

1. INTRODUCTION

In November 2002, the government of Sri Lanka launched *e-Sri Lanka* – the information and communication technology development roadmap to achieve e-governance by the year 2007. Sri Lanka's first ever e-government conference was held in May 2003. The event was given utmost importance by the government of Sri Lanka, and was supported by some of the inter-governmental organizations such as the United States Agency for International Development (USAID) and the Swedish International Development Agency [19].

According to the official document, the main purpose of e-Sri Lanka is to achieve the desired levels of development, by enhancing national competitiveness, reduce or eradicate poverty by realizing enhancements in the quality of life of its citizens [10]. The government of Sri Lanka believes that the vision will

take the dividends of information and communication technology (ICT) to every village, to every citizen, to every business and also transform the way Government works [19].

The main purpose of this paper is to assess e-readiness of Sri Lanka. The uniqueness of this paper lies in the quantitative analysis of the e-readiness of Sri Lanka, whereas many other works engage in qualitative exposure. Based on the assessment, this paper also suggests methods for improving e-readiness of Sri Lanka.

In the next section (section 2), a tool for measuring e-readiness is presented. E-readiness of Sri Lanka is measured in section 3. In section 4, a set of proposals is presented to improve e-readiness and hence e-governance and e-commerce activities in Sri Lanka.

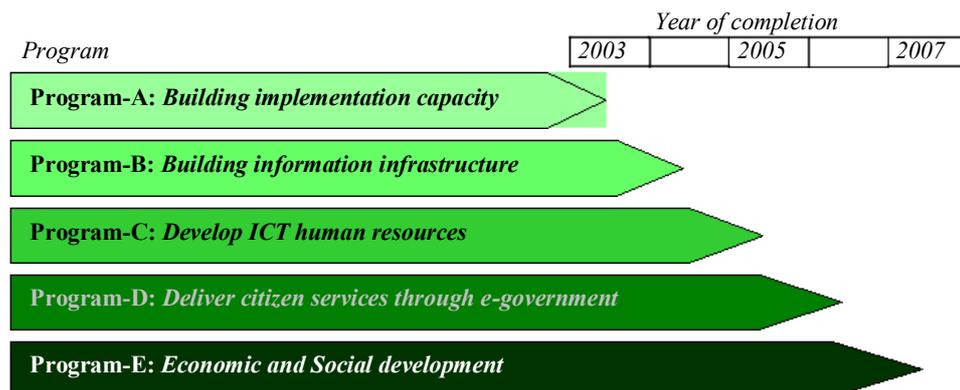


Figure-1: Five-program implementation strategy to achieve the vision of e-Sri Lanka (based on [GoSL, 2003])

2. A TOOL FOR MEASURING E-READINESS

There are many tools available to measure e-readiness of a country. In this work, a tool developed by Bui et al

[2] is used. This tool is selected because it is easily extensible, easy to use and has a large set of indicators.

2.1 The tool

The tool developed by Bui et al [2] consists of three basic building blocks. The three basic building blocks are (figure-1):

- I. Demand forces
- II. Supply forces, and
- III. Societal Infrastructure

The three basic building blocks are divided into eight major factors, and each of these major factors has a set of indicators. The major factors and the number of indicators that come under these factors are given below:

- I: Demand forces:
- i. Culture, understanding and effectiveness: 4 indicators
 - ii. Knowledgeable citizens: 6 indicators
- II: Supply forces:
- i. Industry competitiveness: 7 indicators
 - ii. Access to skilled workforce: 6 indicators
 - iii. Willingness and ability to invest: 4 indicators
- III: Societal Infrastructure:
- i. Cost of living and pricing: 3 indicators
 - ii. Access to advanced infrastructure: 10 indicators
 - iii. Macro economic environment: 12 indicators

The tool uses a total of 52 indicators.

2.2 Measuring e-readiness

All 52 indicators (e_i) are assigned values on a 1-5 scale; 1 is the worst score and 5 is the best score. Then e-readiness of a country is calculated by a simple Figure-of-Merit (FOM) calculation. In this calculation, all the indicator values are multiplied with corresponding weights and summed together.

$$\text{E-readiness} = \frac{\sum_{i=1}^{52} IWeight_i \times e_i}{\sum_{i=1}^{52} IWeight_i} \quad [2]$$

3. MEASURING E-READINESS OF SRI LANKA

In this section, all 52 indicators are evaluated on a 1-5 scale (1 - worst score, 5 – best score).

3.1 Measuring the demand forces

There are two major factors under this block.

3.1.1 Major factor-1: Culture, understanding and effectiveness

Indicator-1. English Language usage: English is commonly used in government and is spoken competently by about 10% of the population [32]. Score: $e_1 = 1.4$

Indicator-2. Percentage of urban population: Urban population in Sri Lanka was 22% in 1998; it was 77% in USA [27]. Score: $e_2 = 2.1$

Indicator-3. Percentage of population over 65 years or older: The percentage for Sri Lanka in 2000 was 6% [20]. Score: $e_3 = 4.9$

Indicator-4. National culture open to foreign influence: N/A

3.1.2 Major factor-2: Knowledgeable citizens

Indicator-5: Adult literacy rate: Sri Lanka has a 91.4% literacy rate [12]. Score: $e_5 = 4.8$

Indicator-6: Secondary enrollment: Sri Lanka has 71% gross secondary enrollment rate in 1998; Norway is one of the leaders under this topic with 117% [28]. Score: $e_6 = 3.4$

Indicator-7: Tertiary enrollment: Tertiary enrollment rate for Sri Lanka is 5%; the rate for Canada is 90% [28]. Score: $e_7 = 1.2$

Indicator-8: MGMT education available in first-class Business Schools: [9] gave 3.7 points to Sri Lanka on a 1-7 scale. Score: $e_8 = 2.8$

Indicator-9: 8th grade achievement in Science: No data available for this indicator. However, going by the diverse material (e.g. [11]) stating that Sri Lanka students perform just about the international average in mathematics, a value of 2.5 is justifiable. Score: $e_{10} = 2.5$

Indicator-10: Flexibility of people to adapt to new challenges: N/A

3.2 Measuring the supply forces

There are three major factors under this block.

3.2.1 Major factor-3: Industry competitiveness

Indicator-11: Technology achievement index: TAI value for Sri Lanka was 0.2 whereas Finland got the highest rating 0.744 [22]. Score: $e_{11} = 2.0$

Indicator-12: Gross tertiary Science & Engineering enrollment ratio: Sri Lanka had 5%. Canada had the highest value for this indicator, 85% [13]. Score: $e_{12} = 1.2$

Indicator-13: Administrative burden for start-ups: On a 1-7 scale, Sri Lanka received 4,9 points [9]. Score: $e_{13} = 3.6$

Indicator-14: Private sector spending on R&D: On a 1-7 scale, [9] awarded 3.10 points to Sri Lanka. Score: $e_{14} = 2.4$

Indicator-15: High-Tech exports as percentage of manufactured exports: Sri Lanka high-tech export is 3% of its total exports. For Singapore, it is 61% [30]. Score: $e_{15} = 1.2$

Indicator-16: Patent applications granted by USPTO: N/A

Indicator-17: Total expenditure for R&D as percentage of GNI: N/A

3.2.2 Major factor-4: Access to skilled workforce

Indicator-18: Public spending on education as percentage of GDP: Citing again [30], Sri Lanka allocated 2.6%; Norway has the highest spending (6.8%) on education as a percentage of GDP. Score: $e_{18} = 2.5$

Indicator-19: Extend of staff training: On a 1-10 scale, [29] gave 2.2 points to Sri Lanka. Score: $e_{19} = 3$

Indicator-20: Research collaboration between companies and universities: On a 1-10 scale, [29] gave 3.0 points to Sri Lanka. Score: $e_{20} = 1.9$

Indicator-21: Number of technical papers per million people: Sweden has the highest papers per million (6.82). For Sri Lanka, it was 1.44 [13]. Score: $e_{21} = 1.8$

Indicator-22: University education meets the needs of economy: Though no valid data is available, according to [11], general literacy for ICT development and growth is insufficient. Score: $e_{22} = 2.0$

Indicator-23: Well-educated people do not emigrate abroad: N/A

3.2.3 Major factor-5: Willingness and ability to invest

Indicator-24: Composite ICRG risk rating: [23] gave 64% (low risk) to Sri Lanka. Score: $e_{24} = 3.6$

Indicator-25: Availability of venture capital: Sri Lanka received 3.1 points on a 1-7 scale [9]. Score: $e_{25} = 2.4$

Indicator-26: Entrepreneurship among managers: [29] awards 5.19 points to Sri Lanka on a 1-10 scale. Score: $e_{26} = 2.8$

Indicator-27: Foreign Direct Investment as percentage of GDP: FDI in Sri Lanka was 1.22%; Singapore has the highest FDI, which is 9.34% of its GDP [20]. Score: $e_{27} = 1.5$

3.3 Measuring the societal infrastructure

There are three major factors under this block.

3.3.1 Major factor-6: Cost of living and pricing

Indicator-28: International cost of living based on US\$ 100: Osaka, Japan is the most expensive place to live (COL index 126.1) and Asuncion, Paraguay is the least expensive place (COL index 36.5). Colombo, Sri Lanka is rated with an index 60 [17]. Score: $e_{28} = 4.0$

Indicator-29: Inflation rate – CPI in percentage: Currently Sri Lanka endures a rate of 8%. Among the best performing countries are USA and Norway both with 1% [3, 20]. Score: $e_{29} = 4.7$

Indicator-30: GDP per capita (PPP) in US\$: PPP for Sri Lanka is US\$ 3279. Norway has one of the highest GDP per capita, currently US\$ 36,000 [30]. Score: $e_{30} = 1.4$

3.3.2 Major factor-7: Access to advanced infrastructure

Indicator-31: Telephone per 1000 people: On a 1-10 scale, Sri Lanka received 4.17 points [30]. Score: $e_{31} = 2.6$

Indicator-32: Mobile phones per 1000 people: On a 1-10 scale, Sri Lanka received 3.17 points [30]. Score: $e_{32} = 2.0$

Indicator-33: Computers per 1000 people: On a 1-10 scale, Sri Lanka received 1.72 points [30]. Score: $e_{33} = 1.3$

Indicator-34: Internet hosts per 10000 people: On a 1-10 scale, Sri Lanka received 0.65 points [30]. Score: $e_{34} = 1.0$

Indicator-35: International telecom, cost of call to US: It cost US\$ 0.30 to make a 5 min call from Norway to USA. From Sri Lanka, the cost is 7.50 [18]. Score: $e_{35} = 1.0$

Indicator-36: E-government: According to [1], USA leads with a score of 3.11 points whereas Sri Lanka got 0.92 points. Score: $e_{36} = 2.2$

Indicator-37: Computer processing power as a % of worldwide MIPS: Sri Lanka's computer processing power is negligible. Score: $e_{37} = 1$

Indicator-38: Freedom on the Internet: [8] gives 52 points out of 100 points. Score: $e_{38} = 2.5$

Indicator-39: Investment in Telecom as a percentage of GDP: N/A

Indicator-40: ICT expenditure as a percentage of GDP: N/A

3.3.3 Major factor-8: Macro economic environment

Indicator-41: Trade as a percentage of GDP: Trade as percentage of GDP should be as low as possible; rate for Sri Lanka is 78 [30]. Score: $e_{41} = 1.8$

Indicator-42: Protection of property rights: On a 1-7 scale, Sri Lanka scored 4.2 points [9]. Score: $e_{42} = 3.1$

Indicator-43: Tariff and non-tariff barriers: On a 1-10 scale, Sri Lanka scored 6 points. Hong Kong and Singapore scored 10 points [29]. Score: $e_{43} = 3.4$

Indicator-44: Soundness of banks: On a 1-7 scale, Sri Lanka received 5.3 points [9]. Score: $e_{44} = 3.9$

Indicator-45: Local competition: On a 1-7 scale, Sri Lanka received 5.10 points [9]. Score: $e_{45} = 3.7$

Indicator-46: Regulatory framework: Singapore is on the top of the list with 1.82 points; Sri Lanka scored 0.38 points in-line with Peru (0.36). Ethiopia was one of the worst performers with -0.71 points [13]. Score: $e_{46} = 2.5$

Indicator-47: Government effectiveness: Singapore is on the top of the list with 2.16 points; Sri Lanka scored -0.44 points among the worst performers like Kazakhstan (-0.61), Maruritania (-0.66) [13]. Score: $e_{47} = 1.3$

Indicator-48: Political stability: Finland is most politically stable (1.61 points). Sri Lanka was awarded -1.63 points, one of the worst performers [13]. Score: $e_{48} = 1.0$

Indicator-49: Press freedom: A press survey [8], places Sri Lanka in the middle of the list. Sri Lanka received 52 points on a 0 (best score) – 100 (worst score) scale. Best performers were Sweden, New Zealand and Iceland, all of these received 8 points; worst performers were Iraq and North Korea, which received 95 and 96 points respectively. Score: $e_{49} = 3.8$

Indicator-50: Rule of law: Singapore is again on the top of the list with 1.85 points; Sri Lanka scored -0.31 points, whereas Nigeria scored -1.13 [13]. Score: $e_{50} = 2.3$

Indicator-51: Control of corruption: Finland is the best performer with 2.25 points; Sri Lanka scored 0.00 points, whereas Nigeria (one of the worst performers) scored -1.05 [13]. Score: $e_{51} = 2.3$

Indicator-52: Adequacy of regulations and supervision of financial institutions: N/A

3.4 The results: the total e-readiness

Assuming equal weights of 1 to all the indicators, we summarize the e-readiness value for each major factor.

Major factor-1: Culture, understanding and effectiveness: $e_{mf_1} = 2.8$

Major factor-2: Knowledgeable citizens: $e_{mf_2} = 2.9$

Major factor-3: Industry competitiveness: $e_{mf_3} = 2.1$

Major factor-4: Access to skilled workforce: $e_{mf_4} = 2.2$
Major factor-5: Willingness and ability to invest: $e_{mf_5} = 2.6$

Major factor-6: Cost of living and pricing: $e_{mf_6} = 3.4$

Major factor-7: Access to advanced infrastructure: $e_{mf_7} = 1.7$

Major factor-8: Macro economic environment: $e_{mf_8} = 2.6$

E-readiness values for each building block is given below:

Basic building block-1: Demand forces:

$$e_{DF} = \frac{e_{mf_1} + e_{mf_2}}{2} = 2.9$$

Basic building block-2: Supply forces:

$$e_{SF} = \frac{e_{mf_3} + e_{mf_4} + e_{mf_5}}{3} = 2.3$$

Basic building block-3: Societal Infrastructure: $e_N =$

$$\frac{e_{mf_6} + e_{mf_7} + e_{mf_8}}{3} = 2.5$$

Summing all these values together:

$$\text{E-Readiness} = \frac{e_{DF} + e_{SF} + e_N}{3} = 2.5$$

4. ANALYSIS

By simply going through indicators, it is easy to find out where Sri Lanka should concentrate to improve its e-readiness.

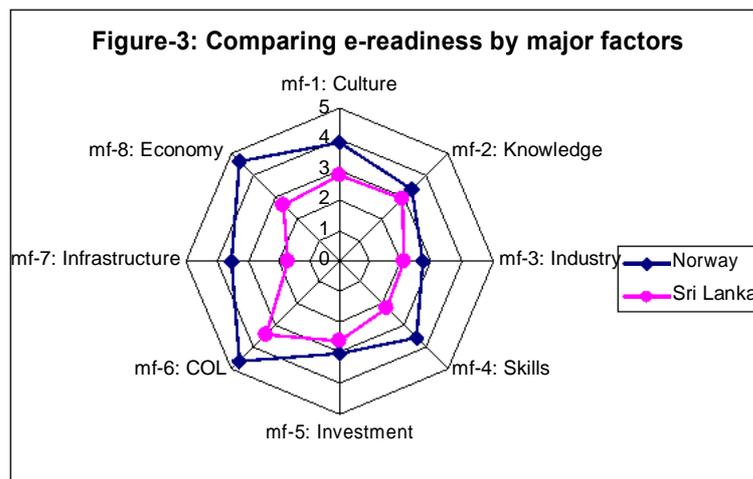
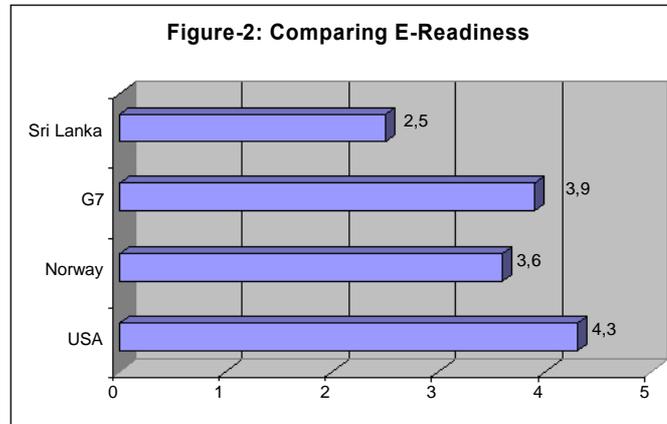
E-readiness values for all the major factors are below average (the average value is 3.0). And hence, the building blocks and the total e-readiness values are all below the average value. Figure-2 plots e-readiness of Sri Lanka against e-readiness of other well-known economies like G7, USA, and Norway; data for Norway is taken from Davidrajah (2004); data for G7 and USA is taken from Bui et al (2001).

Figure-3 depicts detailed benchmarking of e-readiness of Sri Lanka based on the eight major factors. For comparison, values for Norway are also shown in the figure.

In Sri Lanka, demand forces (capability of the people) are about the average value. However, supply forces and societal infrastructure are poor. In some of the

areas (English language usage, tertiary enrollment, high-tech exports, GDP per capita, computers per 1000 people, and telecom costs) Sri Lanka performs poorly. Some other indicators (political stability = 1.0, and government effectiveness = 1.0) show that there is a serious problem in running the country.

Investment in ICT sector is low. After many technology investment debacles, private investors are not so enthusiastic about telecom ventures. The government has also problems in investing in technology sector mainly due to ever increasing health care costs.



REFERENCE

- [1]. ASPA (2001) Benchmarking E-Government: A Global Perspective: <http://unpan1.un.org/intradoc/groups/public/documents/un/unpan003984.pdf>
- [2]. Bui, T., Sebastian, I., Jones, W., and Naklada, S. (2002) E-Commerce Readiness in East Asian APEC Economies, The Pacific Research Institute for Information Systems Management (PRIISM), Honolulu, Hawaii, USA
- [3]. CIAs World Fact book: <http://www.cia.gov/cia/publications/factbook/index.html>
- [4]. Expert (2003): <http://www.expatforum.com/Resources/icol.htm>
- [5]. Reggie Davidrajuh (2004) "Measuring and Improving E-Readiness of Norway", Proceedings of the 2004 International Business Information Management Conference (IBIM'04), Amman - Jordan, July, 2004
- [6]. eEurope (2003) http://europa.eu.int/information_society/europe/index_en.htm
- [7]. Flanagan, T. (2004) Designing a Course on E-Government, Bask Thesis (in Norwegian), Stavanger University College, Norway.
- [8]. Freedom of the Press (2003): <http://www.freedomhouse.org/pfs2003/pfs2003.pdf>
- [9]. Global Competitiveness Report (2001): <http://www.weforum.org/site/homepublic.nsf/Content/Global+Competitiveness+Programme>
- [10]. GoSL (2003) Policy on E-Government, Government of Sri Lanka, May 2003
- [11]. Greenberg ICT Survey (2002) "Country ICT survey for Sri Lanka", Greenberg ICT Survey, Quebec, Canada

- [12]. Human Development Report (2001): www.ssb.no/emner/04/01/notat200306/notat200306.pdf
- [13]. Knowledge Assessment Matrix (2002): http://www1.worldbank.org/gdln-scripts/programs/kam2002/kamscrip.exe/show_page
- [14]. Korea, Ministry of Information and Communication (2003) "Changing Korea with e-Government", <http://www.egov.go.kr>
- [15]. Liu, S. and Hwang, J. (2003) Challenges to transforming IT in the US government, IT Professional, May- June, 2003.
- [16]. McConnell International (2001) "Ready? Net. Go! Partnership Leading the Global Economy", www.McConnellInternational.com
- [17]. Mercer Human Resource Consulting (2003): <http://www.finfacts.ie/costofliving.htm>.
- [18]. Opex Communications (2003): <http://www.opexld.com>
- [19]. SundayLeader (2003) <http://www.thesundayleader.lk/20030504/business-more.htm>
- [20]. SIMA (2003): <http://devdata.worldbank.org/query/>
- [21]. Taylor Nelson Sofres (2002) "Government Online: An International Perspective", <http://www.tnsfres.com/gostudy2002/>
- [22]. Technology Achievement Index (2001): www.undp.org/hdr2001/indicator/indic_267_1_1.html
- [23]. The International Country Risk Guide (2000): <http://www.countrydata.com>
- [24]. Turban, E. And King, D. (2003) Introduction to E-commerce, Prentice-Hall, New Jersey
- [25]. VOX (2003): www.vox.no/prosjektbase/SearchForm.aspx?ECode0=104&EVal0=141&pdid=3761
- [26]. Wong (2000) At the Dawn of E-Government; Deloitte Research, New York
- [27]. World Bank (2000): <http://www.worldbank.org/wdr/2000/pdfs/engtable2.pdf>
- [28]. World Bank (2001): <http://www.worldbank.org/data/wdi2001/index.htm>
- [29]. World Competitiveness Yearbook (2001): <http://www.imd.ch>
- [30]. World Development Indicators (2001): <http://www.worldbank.org/data/wdi2001/index.htm>
- [31]. World Development Indicators (2002): http://www1.worldbank.org/gdln-scripts/programs/kam2002/kamscrip.exe/show_page
- [32]. World Factbook (2003): <http://www.bartleby.com/151/>
- [33]. World Information Technology and Services Alliances (2000) "International Survey on E-commerce", <http://www.witsa.org>
- [34]. Zhou, H. (2001) "Global Perspectives on E-Government", the 3rd Caribbean Ministerial Consultation and High-Level Workshop, December 2001, Jamaica

Internet Role in Lithuania SMEs internationalization

Rimantas Gatautis¹, Panagiotis Damaskopoulos²

¹ Electronic Business Research Center, Kaunas University of Technology, Kaunas, LT 44309, Lithuania

² European Institute of Interdisciplinary Research, Paris, 100084, France
Rimantas.Gatautis@ktu.lt, Takis.Damaskopoulos@eirr.org

ABSTRACT

Small and Medium-sized Enterprises (SMEs) play a crucial role in the economic development of national economies and their growing internationalization. The development of information and communication technologies (ICT) has allowed SMEs to break different barriers, be they economic or regulatory, and to move towards more international positions of doing business. Considering the importance of ICT developments, particularly with regards to international marketing activities and the business opportunities they offer, still few researchers focus on the impact of ICT on international marketing and the internationalization issues it entails. This paper presents findings from a research project which focused upon the use of the Internet by SMEs in Lithuania. It investigates how the level of Internet use in SMEs is related to firm internationalization and Internet marketing activities. The analysis of main factors affecting successful use of ICT such as the role of the Internet in strategies of internationalization, the reduction of barriers through use of the Internet, the influence of the Internet on international marketing activities, and problems of using the Internet for internationalization.

Keywords: *Internet, SMEs, Internationalization*

INTRODUCTION

ICT offers significant opportunities for all enterprises, especially for SMEs, looking to market their products and services internationally. The Internet, for instance, can provide SMEs with a low cost gateway to global markets by helping to overcome many of the barriers to internationalization commonly experienced by small businesses. Such companies are able to compete more easily in the global marketplace, and consumers in emerging markets, in particular, they benefit from expanded range of products, services and information to which the Internet gives them access.

Applying knowledge and using ICT for enterprise internationalization creates the preconditions of development of a knowledge economy. The new opportunities for SME, which can be achieved by using Internet, have become vitally important for US and European enterprises. In Lithuania various aspects cause a problematic situation — lack of knowledge, low level of computerization and Internet usage —with the result that the Internet possibilities for competing in international markets remain largely neglected. The appropriate analysis of using the Internet for internationalization as well as implication for knowledge economy development is a critical task whose solution requires theoretical grounding and research of practical and feasible possibilities.

Problem. *This article is orientated to study the role of Internet for internationalization activities between Lithuania SMEs. The new opportunities for SME, which can be achieved by using Internet, become vitally important for USA and European enterprises. In Lithuania various aspects cause a situation — lack of knowledge, low level of computerization and Internet usage, that the Internet*

possibilities for competing in international markets remain neglected. The appropriate analysis of using internet for internationalization is very important task which solution requires theoretical grounding and research of practical possibilities.

The purpose of research. *The purpose of the research is to analyze the use of Internet between Lithuania's SMEs identifying the internationalization barriers which are removed by Internet, the success factors of using Internet for internationalization and Internet influence on international marketing mix.*

THEORETICAL FRAMEWORK

The Internet offers big opportunities for all enterprises, especially for small and medium-sized enterprises (SMEs), looking to market their products and services internationally. The Internet can provide SMEs with a low cost gateway to global markets by helping to overcome many of the barriers or obstacles to internationalization commonly experienced by small businesses. Such companies are able to compete more easily in the global marketplace, and consumers in emerging markets, in particular, they benefit from expanded range of products, services and information to which the Internet gives them access. However, the problem here is that small businesses have rapidly exploited the new opportunities on their own, but these opportunities also brought many obstacles facing these enterprises.

Firstly, speaking about the internationalization process one has to define what it is and how it affects businesses and especially small and medium-sized enterprises. Internationalization by J.Hamill (1997) is understood as

evolutionary, incremental process whereby SMEs become more and more international over time; the driving force being increased international knowledge, experience and confidence which are seen as breaking down the cultural barriers and reducing the risks associated with further internationalization. Although some important differences exist, the various studies agree that internationalization is a step-wise process which evolves through various stages; for example, the completely uninterested firm, the partially uninterested firm, the exploring firm, the experimental firm, the experienced small exporter, the experienced large exporter.

The main barriers experienced by SMEs when “going global” are psychological, operational, organizational, product/market. At all points one has to understand that the Internet does not provide a panacea for such problems. However, if used effectively, it can be a very powerful tool to assist SMEs in overcoming the main barriers experienced, resulting in more rapid internationalization.

Closely related to the issues of barriers is the question of “critical success factors” in SME internationalization. It is important to observe that there is no such thing as an absolute success factor in the sense of guaranteeing good performance abroad since the factors influencing export success are complex and varied. However, there is sufficient empirical evidence to arrive at some general conclusions concerning the most important factors in successful exporting, such as Finding the “right” overseas agent or distributor, international awareness, knowledge and orientation, etc.

Marketing on the Internet is a very different process from traditional marketing. The international Internet marketing has changed some elements of the marketing mix. The scope of this part was focus on marketing mix dimensions that may be facilitated through the use of the Internet. However, one has to admit that the changes in the four Ps are the result of the interactive nature of the Internet, but this is already the subject of a new international marketing paradigm where the key to a more successful marketing effort on the Internet will be an interactive strategy.

RESEARCH RESULTS

The empirical research has been carried out among Lithuania’s SMEs using a questionnaire on Internet. The point of this survey was to get information preferably from the people who were responsible for the strategic decisions in various enterprises and to acquire a better understanding of the problem. The total number of 147 respondents took part in the survey.

When the respondents were asked for what purposes they used the Internet in foreign markets they had the possibility to mark those answers that best fitted for the

activities of their enterprises. 69% of respondents who answered this question remarked that they were using the Internet for communicating with business partners. As far as this answer was chosen by a good many of respondents it can be stated that primarily the Internet is used for communication with business partners. A little less (67%) account for dissemination of information. Communicating with customers got 60% of the answers, while providing oneself with supplies amounts to very little - only 33% of all answers (See Figure 1).

In conclusion, such results show that for the most part the Internet is used for communication process –either way with business partners and customers; and for the dissemination of information.

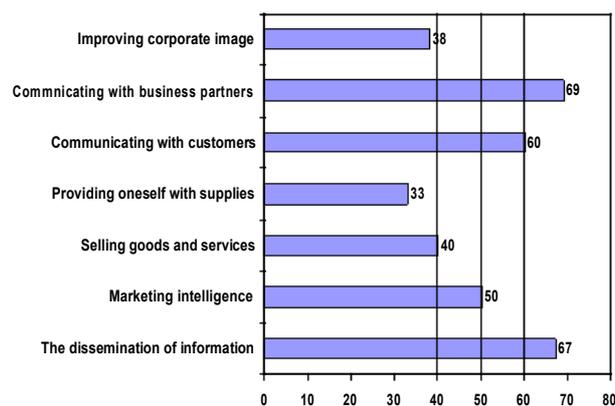


Figure 1 Purposes of using the Internet in the activities of enterprises (%)

The next step in the survey was to sort out what means of the Internet are usually used in the day-to-day activities of enterprises. Answering this question the respondents had to arrange the answers from 1 to 5, where 1 is not used at all and 5 –very important and used much of the time. Hence, the respondents had to evaluate such answers as the Internet website; e-mail; databases; chat rooms, the Internet advertising and the Internet search engines. It can be noted that the majority of respondents (80%) gave 5 points to e-mail which they are using most of the time. 35% of respondents who answered this question stated that using the Internet website in the activities of enterprises could be treated as very important and also gave 5 points to this medium. The same proportion of 24% of the respondents gave to Internet advertising 3 and 4 points which means that using advertising on the Internet is supposed to be less important than using e-mail or the Internet websites. Using databases and the Internet search engines was also evaluated in 5 points respectively by 38% and 27% of respondents. The situation is much different with chat rooms because 76% of respondents gave only 1 point to this medium which means that basically it is not important and not used in enterprises.

Table 1 Means of the Internet most commonly used in the activities of enterprises

	1	2	3	4	5	Response Average
The Internet website	10%	10%	23%	23%	35%	3,6
E-mail	3%	0%	6%	11%	80%	4,7
Data bases	10%	17%	24%	10%	38%	3,5
Chat rooms	76%	16%	4%	0%	4%	1,4
The Internet advertising	14%	17%	24%	24%	21%	3,2
The Internet search engines	15%	15%	18%	24%	27%	3,3

Therefore, these two questions clearly show the trends of the usage of the Internet in the activities of Lithuanian SMEs. The main tool which can embody successfully communication with business partners and customers is e-mail which has been marked by most respondents and surpasses greatly other means like the Internet website; data base; chat rooms; the Internet advertising and the Internet search engines. To show the ratio of the repartition between the means of using the Internet more visibly the response average was calculated and can be seen in Table 1. As it was stated above much importance was attached to e-mail with the average response of 4,7 –the highest from all others. On the subject of chat rooms the response average calculated was only 1,4.

One of the main goals of this survey was to learn more about the importance of obstacles Lithuanian SMEs face in the process of internationalization. The respondents were asked again to evaluate the answers from 1 to 5 where 1 is not important and 5 means that this factor has a large impact on the activities of their enterprises. The results of this question were surprising because the majority of the respondents marked 1 for all the answers. This means that in their estimation there are no obvious barriers for the further internationalization of enterprises in Lithuania. 38% of respondents marked 1 for financial resources while the same factor is very important to only 18% of those who evaluated it with 5 points. 28% of those who answered this question don't feel the lack of ICT at all and 7% of respondents admit that it is a crucial factor for their enterprises. The majority of respondents also gave 1 point to distrust and lack of knowledge in using ICT, correspondingly 28% and 22%. The ones who marked 5 points for these two factors amount to 19% and 16% respectively. To see this striking contrast between matters of substance and factors of no importance only answers which have got points 1 and 5 are presented in Figure 2.

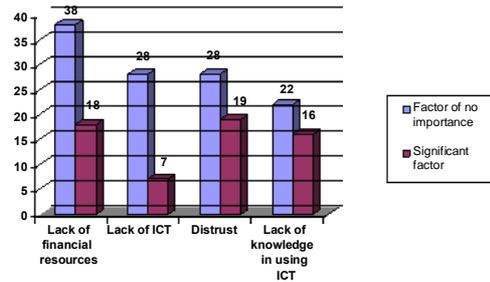


Figure 2 Structure of obstacles affecting the usage of the Internet in the process of internationalization (%)

This figure shows the positive trends towards the process of internationalization and the further development of e-business among Lithuanian SMEs. It is clear that the growing availability of ICT (with prices declining markedly and new models of ICT introduced to the market) has made this factor one of low importance. Nowadays when the development of information society is a top priority more and more models of ICT from low to high prices are available to households, enterprises, national and local government institutions.

Analyzing the results of barriers that the use of the Internet can overcome it is important to find out how the Internet contributes to the diminution of obstacles related to the activities of enterprises when “going global”. 21% of those who responded to the survey stated that the Internet helped their enterprises reduce psychological barriers, which means that through the use of the Internet the enterprises started to think more globally (see Figure 3). They are no longer thinking of exporting as something unrealizable, which is “too risky”“or not for us“. The Internet has helped them overcome initial fears and supported in implementing long rather than short time perspectives. The majority of respondents (64%) admitted that the Internet was of great use in reducing organizational barriers which arise from the lack of limited financial and managerial resources; lack of internationally experienced personnel and their training; lack of knowledge of foreign markets and problems in finding suitable agents and distributors. 57% of respondents declared that usage of the Internet determined the decrease of operational barriers. It became simpler to handle all export documentation through electronic data transfers and make electronic payments. 31% of the respondents noted that the use of the Internet resulted in the reduction of barriers related with product and market. It has become easier to identify and select the most appropriate foreign markets having limited resources for country screening and export research. It is easier to identify the differences of foreign markets and then adapt the products. Only 14% of all who answered noted that using the Internet resulted in having more competitive ability among other enterprises. Nowadays the Internet offers competitive survival rather than a competitive advantage in the

global marketplace. Given that from a competitive standpoint a presence on the Internet is a must in today's business (See Figure 3).

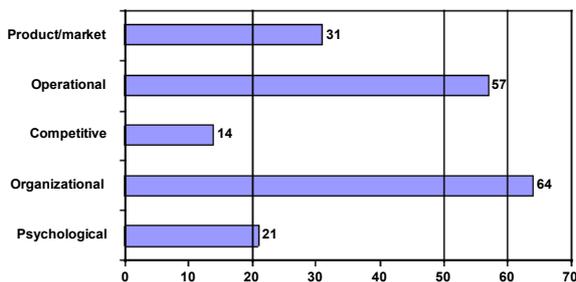


Figure 3 Barriers diminished by using the Internet (%)

It can be concluded that usage of the Internet in Lithuanian SMEs has influenced mostly the reduction of organizational and operational barriers by 64% and 57% respectively in the view of the respondents. But what does this say about the role of the Internet in these processes? How does it contribute to these changes? Firstly, enterprises have access to low cost market research resources. Their knowledge of international markets and culture has improved and at the same time they can reduce their dependence on traditional agents and distributors. The Internet can resolve many of the problems related with the lack of information because much of it can be gathered without any fees or for much smaller ones than in the enterprises of market research.

This questionnaire was formed not only to evaluate the influence of the barriers but also to identify the benefits for the process of internationalization brought by the usage of the Internet. The majority of the respondents who answered (77%) declared that the Internet benefited mostly in the improvement of the communication process. If we look again at the question when the respondents were asked about the purposes of using the Internet in foreign markets, the Internet was used mainly for communication with business partners. Once again the conclusion may be drawn that the Internet is the source of communication that is facilitated greatly through the use of this medium. The same percentage of respondents (23%) gave 1, 2 and 4 points to the first answer about finding new markets with the help of the Internet. Such results show that the Internet hasn't helped much in finding new markets for almost half of the respondents. The situation is different with finding new customers because 50% of respondents gave 4 and 5 points to this answer. In sum, the majority of the respondents believe that the Internet is more useful in finding new customers than new markets. The biggest share of the respondents also acknowledged that the Internet reduced the costs of their activities and gave quick access to all the information. These two answers were evaluated at the maximum with 5 points, respectively by 38% and 57% of the respondents. To see the results more clearly the response

average again was calculated. The figure of 4,6 confirms once again that the Internet is the most effective medium for the process of communication. The reduction of financial costs is evaluated with the average of 3,8 which means that the impact of the Internet on communication process and quick access to the information is bigger than that on financial costs. The response average of 2,9 (the lowest from all others) implies that the Internet is less useful in finding new markets than in any other areas.

Table 2 Benefits originated from the use of the Internet

	1	2	3	4	5	Response Average
Helped to find new markets	23% (7)	23% (7)	13% (4)	23% (7)	17% (5)	2,9
Was useful in finding new customers	4% (1)	25% (7)	2% (6)	25% (7)	25% (7)	3,4
Contributed to the improved communication process	6% (2)	0% (0)	3% (1)	14% (5)	77% (27)	4,6
Made the costs of the activities lower	9% (3)	12% (4)	12% (4)	29% (10)	38% (13)	3,8
Gave quick access to all the information necessary	6% (2)	0% (0)	11% (4)	26% (9)	57% (20)	4,3

The respondents were asked which factors they considered to be crucial for successful usage of the Internet in the internationalization process by Lithuanian SMEs. The most critical factor given by the survey respondents has to do with knowledge and skills (see Figure 4). The next two factors are financial resources and new technologies. The respondents had to choose among the factors, which helped to use the Internet successfully in the process of internationalization. Most of the respondents believe that knowledge and the ability to use the skills properly is a determining factor in the successful use of the Internet. When calculating the response average, this factor got 3,8. New technologies with the average figure of 3,4 is the second important factor and financial resources with the average of 3,2 is supposed to be less important than knowledge and new technologies. Figure 4 shows the percentage of respondents who gave 1 point to unimportant factors and 5 points to significant factors.

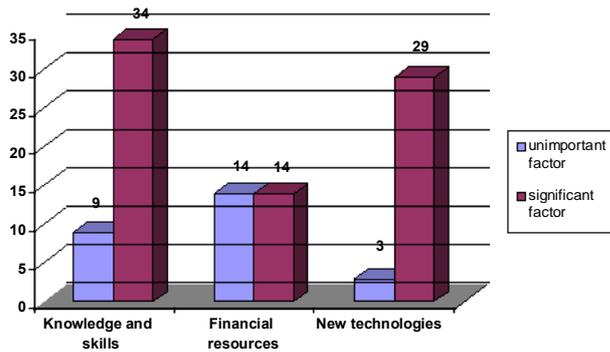


Figure 4. Structure of factors which contribute to the successful use of the Internet (%)

In summary, it can be stated that for Lithuanian SMEs knowledge and skills (human resources on the whole) is likely to play a key role in the industry base of technology and knowledge-intensive economic development. It is clear that these factors cannot be separated from one another because in order to develop the skills and acquire relevant knowledge one has to have some financial resources to purchase a computer and have the Internet access. The point to be stressed here is simply that support of the educational sector is a crucial condition for a sustainable transition to a knowledge economy. It is desirable that pupils and students have a higher level of computerization and the Internet access in educational institutions where they develop their knowledge and skills to become highly skilled ICT manpower. Such people are the backbone of electronic business development in Lithuania and further process of internationalization in enterprises. One of the most interesting aspects of the process of internationalization is the effect of the Internet on the changing face of the marketing mix. According to various authors the Internet has created a new environment for international marketing and similarly the changes of this environment resulted in the changes of 4 Ps (Product, Price, Place, and Promotion). The Internet has often led to faster discovery of customer needs; increasing competition and standardization of prices; interactive communication with customers and reduced importance of traditional intermediaries.

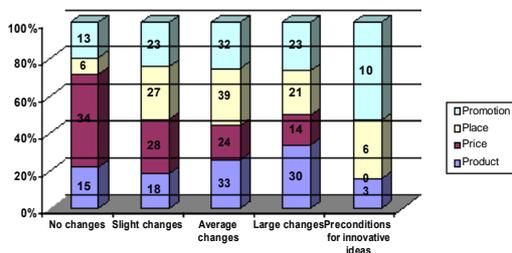


Figure 5 The impact of the Internet on marketing mix (%)

It was necessary to identify the importance of these changes of marketing mix in the activities of Lithuanian

enterprises. 33% of our survey respondents stated that the usage of the Internet resulted in the average changes of the product or service (see Figure 5). There were average changes in customization of the product, discovery of customer needs and product testing in the biggest part of Lithuanian enterprises. The biggest share of the respondents about price (34%) declared that there were no changes in prices in their activities in foreign markets. 39% of respondents experienced average changes in distribution channel and 32% of respondents admitted that the use of the Internet has also caused average changes in promotion.

These figures indicate that the Internet has average influence on almost all of the components of marketing mix except price. The majority of Lithuanian SMEs don't experience any changes of prices when "going global". Figure 5 shows the trends more clearly. Speaking about promotion, 10% of enterprises believe that the Internet is the source of innovative ideas. This could be explained by the nature of online promotion techniques, especially advertising which is perhaps the best developed use of the promotion can be implemented in various ways and there is always space for creative ideas. From this standpoint the Internet is understood as a mean which can quicken generation and realization of the ideas. One has to understand that features valuable in promotion don't fit at all to the process of fixing prices where thoroughness and accuracy but not creativity is valuable. Maybe it is for this reason that use of the Internet has less power to change prices in enterprises.

CONCLUSIONS

Summarizing the research results it can be stated that the largest part of Lithuanian enterprises (67%) use the Internet for communication with their national and foreign partners. The main purposes of using the Internet are primarily communication with business partners, customers and dissemination of information. The Internet is used least while providing oneself with supplies. The most popular and widely used tool of the Internet is e-mail. It surpasses greatly other means of the Internet like the Internet websites; data bases; the Internet advertising and search engines. These tools are used much less while chat rooms are not popular at all in the activities of Lithuanian enterprises.

The structure of obstacles affecting the usage of the Internet shows the positive trends towards the process of internationalization. This may be explained by the fact that the majority of Lithuanian SMEs do not consider distrust; lack of financial resources, lack of communication technologies and knowledge as significant obstacles for internationalization.

The survey also shows that the Internet reduces organizational barriers more than any others. As well Internet plays important role in reducing operational and

product/market barriers.

The main benefits of Internet use in the day-to-day activities of Lithuanian SMEs result in improved communication process and quick access to all the information necessary.

The prior factor which contributes to the successful use of the Internet in the process of internationalization is knowledge and skills. Financial resources and new technologies are ranked as less important.

The Internet has average impact on three elements of the marketing mix: product; place and promotion. However, the majority of Lithuanian enterprises do not feel any changes of prices when operating in global environment.

Consequently, 77% of enterprises believe that the Internet use is essential and has large impact while operating in foreign markets and only 6% of enterprises assume that the Internet has no power in the process of internationalization.

Unless the technological backbone is easily accessible for internationalization activities, the lack of knowledge for using technologies prevents companies from productive use of technologies. Nevertheless the importance of Internet in competing in global market is recognized by many businessmen's.

REFERENCES

- [1] Eid, R. and M. Trueman, "The Internet: new international marketing issues". *Management research news*, Vol. 25 No. 12, pp. 54-67, 2002.
- [2] Hamill, J. "The Internet and international marketing". *International Marketing Review*, Vol. 14 No. 5, pp. 300-323, 1997.
- [3] Information and communications technology // Country profile: Lithuania. International trade centre, UNCTAD/WTO. Ženeva, 2002.
- [4] Johanson J., J.E. Vahlne "The internationalization process of the firm: a model of knowledge development and increasing foreign market commitments". *Journal of International business studies*, 8 (1), pp. 23-32., 1997.
- [5] Lithuanian National Information Society Development Conception. Lithuanian Government, 2001.
- [6] Oystein M. "The internet and international marketing: An Empirical Analysis of Small and Medium Sized Norwegian Firms". *Quarterly journal of electronic commerce*. Vol. 3/No. 1, pp. 31-41., 2002.
- [7] Palumbo, F. , P. Herbig., "International marketing tool: the Internet". *Industrial Management&Data systems*, No. 6, pp. 253-261., 1998.
- [8] Quelch, J.A., L.R. Klein, "The Internet and international marketing". *Sloan Management Review*, Spring, pp. 60-75., 1996.
- [9] Samiee, S., "The Internet and international marketing: is there a fit?". *Journal of interactive marketing*, Vol. 12 No. 4, Autumn, pp. 5-21, 1998
- [10] Survey report "Digital Lithuania". Vilnius, VU, 2001.
- [11] Survey reports "IT and Internet marketing". SIC market research, 2001, 2002, 2003.
- [12] The Knowledge Economy in Lithuania. A study of the industry's prospects // Ekonomines konsultacijos ir tyrimai. Vilnius, 2003.
- [13] Zilioniene, I., "Why we fear Digital Divide". *BALTIC IT&T REVIEW*, No. 4 (31), 2003.

Investigating the Use of the Business Model Concept through Interviews

Alexander Osterwalder, Yves Pigneur

HEC Business School, University of Lausanne, Switzerland
{Alexander.Osterwalder, Yves.Pigneur}@unil.ch

ABSTRACT

The term *business model* is widely used and is increasingly gaining importance as a business concept and management tool, particularly in e-business. Several authors propose the application of the business model concept in different domains, such as for e-business planning, IS alignment, visualization, innovation, e-business process design or improving communication between stakeholders. Yet, little has been said on the business world's perception of the use of business models, which this paper attempts to investigate through semi-structured interviews with company executives and business consultants.

Keywords: Business Models, Business Model Concept, E-Business, Management, Interviews, Ontology

1. INTRODUCTION

The term and concept of business models is continuously gaining in acceptance and importance after overcoming its passage as a buzzword. A query in Business Source Premier, a leading electronic database for scholarly business journals, shows that in the 1990s a veritable boom of the expression business model has taken place (see Figure 1). The term is found in numerous variations, such as “new business models”, “e-business models” or “internet business models” and has occurred 144 times in an abstract and 29 times in the title of a peer-reviewed article in the above-mentioned database.

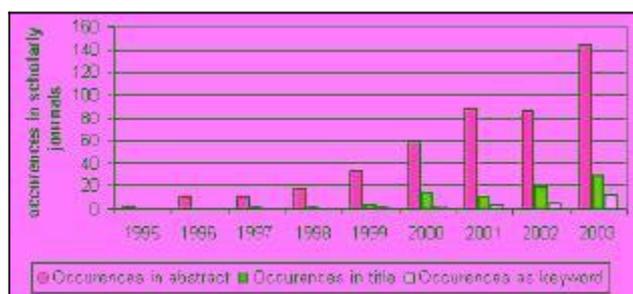


Figure 1 : Occurrences of the term business model in Business Source Premier

Yet, far more interesting than the number of occurrences of the term is the appearance of the business model concept as a management tool as outlined in a number of literature reviews by Stähler [12], Pateli and Giaglis [9], Rentmeister and Klein [10] and Osterwalder [6]. Most promisingly, several authors propose a conceptual approach to business models by describing and modelling a business model's elements and their relationships [1, 2, 4, 6, 13]. Furthermore, they outline a number of possible applications for such a conceptual approach, like business planning, IS alignment, visualization, innovation, process design or improving communication.

This paper aims at taking up the use of conceptual business models by investigating how the business world perceives the use of business models as a management tool. So far, little business model research has attempted to describe the confrontation between concepts and business world, with the exception maybe of Linder [5] and Gordijn [3]. The former has conducted interviews with company executives investigating business model relevance while the latter has applied the e³-value business model framework in consultancy work and published the experiences.

The originality of this paper is that it explores the business world's interest in and perception of different uses of the business model concept. In a dozen semi-structured interviews we captured the opinion of business leaders and consultant and subsequently analyzed the interview transcripts.

The paper is structured as follows. In the next section we explain the business model ontology that has served as the basis of the interviews on the use of business models. The third section outlines the methodology. In section four we describe the interview outcome and analysis before concluding in section five.

2. THE BUSINESS MODEL ONTOLOGY

To explain the interviewees the concept of business models we used Osterwalder and Pigneur's business model ontology as it represents a consolidation and synthesis of ongoing research in the business model domain [7, 8].

The business model ontology captures the money earning logic of a company and consists of nine main components. These so-called business model building blocks are *value proposition*, *customer relationship*, *distribution channel*, *target customer*, *value configuration*, *capability*, *partnership*, *revenue model* and *cost structure* (see Figure 2).

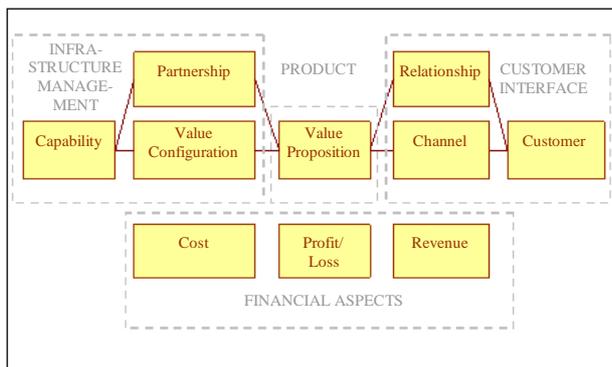


Figure 2 : The Business Model Ontology [6]

The starting point for investigating business model use in the interviews was the presentation of easyJet.com's business model (see Figure 3). The business model building blocks of this bird's eye view were also presented in more detail and in more fine-grained examples.

	PARTNERSHIP		easyJet Business Model (bird's view)		RELATIONSHIP
	Regional airports UK air traffic control system (NATS)				Personalized e-mail relation Strong easyJet brand name
CAPABILITIES	VALUE CONFIGURATION	VALUE PROPOSITION	CHANNELS	TARGET CUSTOMERS	
Full airplanes (80-85%) Low turnaround times of airplanes	Value Network (Stabel, 1998)	Low fares (44 routes/17 cities) Easy online booking Transparent	Direct sales only Internet (90%) Call center (10%)	Low fare segment (40%) Private (40%) Business (60%)	
	COST MODEL	PROFIT/LOSS	REVENUE MODEL		
	Low cost through online sales, ticketless flights, no free lunch, cheap airports, paperless office	Turnover 356 mio £ Profit 40 mio £	Yield management (dynamically forecasted prices based on offer and		

Figure 3: easyJet.com business model

3. METHODOLOGY

Between June and October 2003 we conducted a dozen 60 to 90 minutes long interviews with managers and consultants to assess the use of business models in the firm. These were a series of semi-structured interviews that aimed at investigating the relevance of business model research exploring possible uses of the business model concept. The interviews were all transcribed and subsequently compared and analyzed.

We introduced the interviewees to the business model concept by explaining the business model ontology described in section two. The interview itself then focused on the business model concept and its use in general.

The interview was structured into four parts (see Table 1). A first part was composed of questions on the use of models, formal models and business tools in the

interviewees' companies. The second part consisted of a short explanation of the business model ontology by presenting a number of examples of business models or parts of a business model, such as easyJet.com, ColorMailer, Barnes&Noble and Nokia. The third part was destined at discussing Osterwalder and Pigneur's business model ontology. The goal of the most extensive fourth part was to investigate in which fields the business model concept could make a contribution.

Table 1: Business Practitioner Interview Structure

Question Domain	Questions
Questions on the use of business concepts & tools	How do you plan the general business objectives of your company? Do you use any conceptual tools to plan your business or to sketch the general direction in which your firm is heading?
	If yes, do you use any specific formalism(s) to do this?
	If yes, do you use any specific software tool to do this? If yes, which one(s)?
Demonstrations & Explanation of the Ontology	easyJet.com, ColorMailer, Barnes & Noble, Nokia
Questions on the fidelity with real word phenomena	In your opinion, what elements are missing in the model presented before?
	In your opinion, what elements should not belong to the model presented before?
	How could such a model help you define business indicators?
	How could such a model help you or a group of managers make better decisions?
	How could such a model improve some parts of strategic planning?
	How could such a model make it easier to chose and design appropriate information systems (e.g. software purchases like Customer Relationship Management or Supply Chain Management...)
	How would it be able to foster innovation in a company with such a model?
	How do you think such a model could improve business process design and engineering?
	How could it be helpful to have such a model to communicate your business model. (when making decisions, to communicate with employees)
Final discussion	Do you have any final comments?

The interviewees were conducted with 8 managers and 3 consultants. The managers represented companies in the domains of transport (1), media (1), services (4), retail (1) and industry (1) (cf. Table 2). All enterprises except the industrial company had a strong Internet component. Company sizes ranged from 3 to 3'315 employees. From the companies addressed and asked to participate in the interviews only one did not take part after initially agreeing. A general impression was that the interviewees were all quite receptive to the topic of business models. One manager of a start-up company mentioned: "I'm already happy that somebody tries to define the term business model. It was one of the most violated terms. Everything was a business model. Everybody asked me what a business model is. I could never really define it. It is good that somebody is looking at this".

4. INTERVIEW OUTCOME AND ANALYSIS

The first part of the interview aimed at examining if the interviewees and their companies used models, formal models or even more sophisticated business tools in business planning. Of course this was not aimed at giving us such detailed insights to business concept use, as for example the study by Rigby [11], but it can serve as an indicator for companies being comfortable with the use of concepts. Not surprisingly, very few companies, but all the consultants used concepts or models. One consultant put it very clearly: *“I have seen very few managers use concepts or models, but then that*

is one of the reasons why they bring in us consultants”. The CEO of the industrial company that actually used some formal concepts said that *“it took quite a lot of work to use these concepts, but it was very helpful. What was not satisfying was that we didn’t exploit them sufficiently once the work was done”*. At another point he mentioned that *“at the end of the day there stays very little time for doing prospective business planning”*. Surprisingly, the smallest company interviewed used a conceptual tool for business planning. In the CEO’s opinion the most important aspect of using a concept was consistency. Concepts would only bring advantages when rigorously applied.

Table 2: Interview outcome

Retail over Internet	3	little	no	yes	q1		q2	q3	q4	q5	
Software in the mobile industry	5	little	no	yes		q6	q7	q8	q9		q10
Service over Internet	15	no	no	yes	q11	q12	q13		q14	q15	
Electronic trading platform in Finance	31	no	no	no		q16		q17			
Internet Industry Platform	80	yes	no	yes		q18		q19	q20	q21	
Industry	400	yes	no	no		q22		q23	q24	q25	q26
Entertainment with Internet component	10-1200	no	no	yes			q27				
Transport with Internet component	3'315	no	no	yes		q28	q29	q30	q31		
Consultant 1		yes	no	-	q32	q33, q34		q35	q36	q37	q38
Consultant 2		yes	no	-	q39	q40	q41		q42		q43
Consultant 3		yes	no	-		q44					

■ positive answers
 ■ negative answers
 ■ neutral answers
 q = quote

The younger companies had all made business plans during their start-up phase. This could also be seen as the use of some sort of concept, as most of them had structured their business plan according to recommendations in books available on the market. One executive compared the business model concept to the business plan approach and saw it as a quite useful checklist, particularly for start-ups.

A surprising finding was that a large number of companies relied deeply on trial and error to introduce new aspects to their business. The planning for this relied essentially on drafting text documents with a Word processor and calculating the financials with a spreadsheet program.

The fourth and major part of the interview aimed at investigating the interviewees' opinion on different business model concept applications. Seven areas were

identified and presented to the interviewees, namely the definition of indicators, decision making, strategic planning, designing information systems, innovation, process design and improved communication.

The two most important themes that were isolated from the answers to the seven questions were transparency and communication. They appeared several times as answers to several different questions. Transparency was particularly mentioned by two managers and two consultants. For example to the first question on business indicators one manager mentioned: *“it is important to make things transparent to show where cost and risks come from. Transparency is very important. [...] Such a framework can be interesting”* (cf. quote q1 in Table 2).

Communication was another recurring theme throughout the answers and explicitly addressed in the last question on communication. Especially in question

two on decision making communication was mentioned by four interviewees in combination with transparency: *"it's also about transparency. Somebody puts up the sheet and says this is our company and this is the way it works. Not everybody understands the same thing under the functioning of a company"* (q33). One manager that was less enthusiastic about the business model concept's role in improving decision making acknowledged: *"it helps in communicating, in that sense it may improve decision making. But at least this way people talk about the same thing"* (q12). One CEO of an industrial company said that a business model *"can be very helpful if not too complicated and adapted to specific managerial levels for decision making: employee communication; customer presentations (screened); and training purposes (employees and customers)"* (q26) and that *"it [the business model] is a very useful instrument to initiate discussions with employees, partners or customers about process improvements. The visualization helps also in the internal or external communication of business decisions"* (q25).

Table 2 gives an overview of the interviews depicted in the following sections of the paper, showing the interviewee's economic domain, the number of employees, the company's use of concepts and tools, as well as the use of trial & error, and, the interviewee's position concerning the questions. As outlined above the questions ranged from *"can the business model concept help in defining indicators"* to *"can the business model concept improve communication"*. Favourable answers are indicated in green, reluctant ones in red and neutral ones in grey.

4.1 Defining Indicators

The first question concerned defining indicators in a rapidly changing business environment characterized by ICTs and e-business. While quite a few interviewees seemed favourable to the idea that the use of the business model concept could improve defining business indicators they particularly stressed financial indicators. One interviewee pointed out the importance of transparency regarding costs and risks (q1). A consultant insisted on the financial relationships: *"If I can change parameters in the model and see how this impacts costs, profits, then yes. If you can model cost structure and profitability you have to integrate it with the how and who [infrastructure and customer relationship]. I have to understand how factors influence each other and then I will look at evolutions more closely with my controllers"* (q32). The chief operating officer (COO) of a company with 15 employees was rather critical and remarked that *"in a company like ours everything is in the heads of the people"*. He saw the value of the business model approach elsewhere: *"Where it can be very important is on the venture capital side or the investor side where you have to map business models quickly"* (q11).

Astonishingly, few interviewees bought in on the idea of not only defining financial indicators, but defining indicators throughout the business model pillars. One consultant that understood this asked in what the idea of deriving indicators from the business model was different from the balance scorecard concept (q39).

4.2 Improving Decision making

The reactions to the second question on improved decision making through tools based on the business model approach were mixed. Many of the interviewed saw an indirect influence on decision making through discussing the fundamentals of a business (q34, q40, q18, q28) and through improving communication. The executive from the transport industry saw the force of schematizing in asking fundamental questions: *"You will have to ask yourself in which box [element] I put what - it allows me to look at things from a new angle. It forces you to ask questions. But I see this rather as a tool for start-ups"* (q28). One manager of a start-up company talked about his experience at a large corporation where they used a simple conceptual model to communicate during meetings. As regards the business model concept he noticed that it may help people talk about the same thing and thus improve decision making (q12). Another COO was quite enthusiastic stating *"it would certainly help to make better decisions. Oh yeah, I'd love to see this in my department"* (q16). One of the interviewed consultants acknowledged the value of the business model concept in communication but perceived it as too static to improve decision making (q44). The founder of a software company in the mobile industry felt that the problems he experienced in decision making where much more related to human aspects of the deciders (q6). The CEO of an industrial company was rather favourable, but questioned the availability of business data, *"the framework can help to better set priorities, however, to better support the decision process, the availability of business data needs to be ensured"* (q22). Furthermore, he stressed the importance of having to be able to *"introduce such a business model framework within reasonable means and resources"*.

4.3 Improving strategic planning

The business model concept's possible direct or indirect contribution to improving strategic planning was perceived by all but two interviewees that answered the question. The CEO of the smallest company interviewed saw the business model concept as a way to outline the steps necessary to achieve a strategic goal (q2). Another manager saw the advantage in taking into account all the elements of a business model (q7). A consultant proposed combining business model and scenario approach in order to have an impact on strategic planning: *"[The business model concept] can indirectly influence direct strategic planning through scenarios and transparency"* (q41). One executive that was

reluctant stated: *"We would think what kind of skills we need. We would make a cost model and a revenue model. We would make a prototype and look if it works. If it doesn't work we would stop the test"* (q13). This approach remarkably resembles some of the structures of the business model concept. The executive from the transport industry was also reluctant and thought that the ontology could have an impact on strategic planning if it allowed simulation (q29). Finally, the executive from the entertainment industry observed that looking at his enterprise through the business model lens opened up new perspectives that were not usual to his business (q27).

4.4 Helping in the Design of ISs

The fourth question on improving the design of information systems (IS) was not answered by all the interviewed practitioners because not all of them felt expert enough to reply. The answers that were given were quite mixed and split between three supporters and three opponents of the idea. Noteworthy, the supporters felt very strong about the necessity to describe the business logic of a company to improve IS design. The CEO of the industrial company, the COO of the financial platform and one of the management consultants strongly backed the link between business model and IS design. The first declared *"once business models/processes are clearly defined and evaluated, respective appropriate information systems can be easily defined. There are significant differences in the business processes supported by off-the-shelf information systems. Therefore, a business framework can not only help, but is almost required to identify the information system with the best fit"* (q23). The second proponent mentioned: *"I think you make things much easier here... much, much easier. Well I think what we've done is that [...] it presents it [the business model] in a clear way. So everybody starts in the same place. Picking up on your point about taking it out of a manager's head and putting it on a piece of paper"* (q17). The consultant answered like the CEO: *"You absolutely need a model before you can do anything"* and *"especially if the dependencies and interactions of the different building blocks become clearer. Transparency is important for understanding what we do and how the building blocks interact"* (q35). The CEO of the small Internet retailer felt that the business model concept would probably not improve IS design because of its static nature (q3). Similarly, the co-founder of the small software company remarked that the tools based on the business model approach *"may influence IS design, but will not essentially help to improve it"* (q8). An executive from an Internet industry platform had the impression that according to his experiences *"engineers would not be favourable to use such a tool as the business model concept"* (q19). Another manager saw the use of the ontology in IS design particularly for new economy companies of a networked type (q30).

4.5 Increasing innovation

With regard to innovation there were two positions, the first accentuated the human capability aspect of innovation (q9, q14, q42) and the second drew attention to the fact that the visualization and transparency created by business models could improve innovation (q4, q36, q31). A proponent of the first view stated: *"What is important is thinking it through. That is what takes time. If we look at the value proposition it is the thinking behind it that is important. If we are in front of investors the important thing is that it is innovative, that there's a prototype, that it was tested and that it comes over technically. And that thinking takes time"* (q14). Similarly, another one mentioned that *"innovation is more about the people"* (q9). A consultant commented that the business model approach may help in innovating but that *"the tool will not generate innovation. You still need the creativity of the people"* (q42). A proponent of the positive effect on innovativeness pointed out that thinking through a business model will automatically stimulate innovation: *"Establishing/elaborating business models and business processes requires thorough and in-depth analyses which leads by itself to innovative ideas and incentives on how to improve key indicators and elements of business models and how to better process steps or phases"*, (q24). Similarly, another executive had the impression that structuring the thought process could be an advantage for brainstorming and have an effect on innovation (q20).

4.6 Improving Process Design

The interviewees that answered the question on improving process design were all relatively positive except for one that underlined the model's static nature as he had already done for information systems (q5). They saw the business model's value in its high level business logic description: *"I think something like this model is the foundation. Before you can describe processes you have to have something like this at the generic high level"* (q37). The executive of a start-up, however, emphasized that this is too sophisticated for small companies, *"in a small company this would be overkill"* (q15). The executive from the Internet industry platform had the impression that such a model would help people keep in mind the whole, because they would often forget other parts of the company when designing processes (q21).

4.7 Improving Communication

As explained above its ability to improve communication between different parties was seen as one of the strong points of the business model concept. One consultant saw it as *"a formalism that everybody understands to speak the same language"* (q43). Another one pointed out that it helped people to talk

about the same thing: "First you have to assure that everybody is talking about the same thing. For example in customer relationships you can imagine talking about hundred different things. It's about presenting things in a simple way that is certainly the most important" (q38). One interviewee saw tools based on the business model concept as a means to communicate with employees, customers and for training purposes (q26). The interviewee opposed to the positive effect on communication felt that "it's too complicated to represent a business model in 9 elements. It's [a business model] more about how I make my money" (q10).

5. CONCLUSION

Recapitulating, the interviews have shown a number of different things. The executives of very small companies have mixed opinions concerning the use of the business model concept, particularly in applying it to their firm. Though not completely reluctant to the use of the business model concept they feel that the business model of their company is already sufficiently clear. It also seems that the larger the companies of the interviewees were the more favourable they tended to be regarding the use of the business model concept - with the exception of the executive from the transport industry. The consultants all seemed quite comfortable with the idea of applying the business model concept to a number of different problem domains. The main strengths of the business model concept that came out of the interviews were threefold:

- a) The ability to create a transparent big picture of a business and to externalize the relationships and dependencies of the business elements.
- b) Its use is comparable to the use of a commonly understood language to enable communication.
- c) It can stimulate approaching and understanding the fundamental questions of a business.

Of course the above mentioned strengths are simply impressions of the interviewed business practitioners. Yet, they give a good impression of how business practitioners feel of the use of the business model concept in different applications. However, further research would include applying different business

model approaches to different concrete cases and application domains in order to assess their performance in the field.

REFERENCES

- [1] Afuah, A. and C. Tucci, *Internet Business Models and Strategies*. 2nd edition ed. 2003, Boston: McGraw Hill.
- [2] Gordijn, J., Value-based Requirements Engineering - Exploring Innovative e-Commerce Ideas. 2002, Vrije Universiteit: Amsterdam, NL.
- [3] Gordijn, J. and J.M. Akkermans. Does e-Business Modeling Really Help? in 36th Hawaii International Conference On System Sciences. 2003. Hawaii: IEEE.
- [4] Hamel, G., *Leading the revolution*. 2000, Boston: Harvard Business School Press.
- [5] Linder, J. and S. Cantrell, *Changing Business Models: Surveying the Landscape*. 2000, accenture Institute for Strategic Change.
- [6] Osterwalder, A., *The Business Model Ontology - a proposition in a design science approach*, in *Institut d'Informatique et Organisation*. 2004, University of Lausanne, Ecole des Hautes Etudes Commerciales HEC: Lausanne, Switzerland. p. 173.
- [7] Osterwalder, A. and Y. Pigneur. An e-Business Model Ontology for Modeling e-Business. in 15th Bled Electronic Commerce Conference. 2002. Bled, Slovenia.
- [8] Osterwalder, A. and Y. Pigneur, *An ontology for e-business models*, in *Value Creation from E-Business Models*, W. Currie, Editor. 2004, Butterworth-Heinemann.
- [9] Pateli, A. and G. Giaglis. A Framework For Understanding and Analysing e-Business Models. in Bled Electronic Commerce Conference 2003. 2003.
- [10] Rentmeister, J. and S. Klein, *Geschäftsmodelle - ein Modebegriff auf der Waagschale*. Zeitschrift für Betriebswirtschaft, 2003(73): p. 17-30.
- [11] Rigby, D., *Management tools and techniques: A survey*. California Management Review, 2001. 43(2): p. 139-+.
- [12] Stähler, P., Geschäftsmodelle in der digitalen Ökonomie. Merkmale, Strategien und Auswirkungen, in MCM. 2001, University of St.Gallen HSG: St.Gallen.
- [13] Tapscott, D., D. Ticoll, and A. Lowy, *Digital Capital - Harnessing the Power of Business Webs*. 2000, Boston: Harvard Business School Press.

Logistical Performance as a Key Factor of Success for European Food E-tailers— Three Paradoxes and a Research Agenda

Dominique Bonet¹, Gilles Paché²

¹ CRET-LOG, Université de la Méditerranée (Aix-Marseille II), France
dbonet@univ-aix.fr

² ERFI-CEROM, Université Montpellier I, France
gpache@univ-montpl.fr

ABSTRACT

In many Northern European countries, the sale of convenience goods through the Internet is experiencing a slow but undeniable growth. The emergence of this new distribution channel has brought into existence an extremely productive research area on the key factors which food e-tailers need to master for a lasting success in this field. The central nature of a perfect logistical organization is particularly emphasized in the areas of cost and service quality, both in terms of order fulfillment and order delivery. There is no question of minimizing the importance of logistics in the success of online transactions, as e-consumers obviously expect to be delivered at home in good conditions. But can we affirm that logistical performance systematically plays a crucial role for food e-tailers in obtaining a sustainable competitive advantage? By critically examining the relevant literature, this paper attempts to prove that the answer may be negative. The authors refer to three paradoxes relating to e-grocery logistics: one, the pressure of national and local Authorities leads to the sharing of resources to reduce environmental pollution (first paradox); two, the recognition of an e-marketing mix dissolves logistical performance in a set of non-logistical variables whose weight is probably significant in e-consumers' behavior (second and third paradoxes).

Keywords: consumer expectations, e-grocery, e-marketing mix, Internet shopping, logistics, order delivery, order fulfillment, package deal of services

1. INTRODUCTION

The sale of convenience goods by the Internet, known under the generic term of e-grocery, is experiencing an appreciable development in many European countries, even if the online channel still remains marginal compared to the traditional in-store channel. The act of physical shopping is still often “*a socially valuable and personally relaxing experience*” [12, p.231], but a new generation of highly educated and professionally hyper-active young consumers find an increasing interest in online transactions for convenience goods which do not create any “retail therapy” buzz of in-store shopping. Numerous works now focus on e-grocery, systematically emphasizing the importance of logistics—the order fulfillment and order delivery processes—as key factors of success for e-grocers [8]. Of course, delivering the right product at the right time at an acceptable cost is an important stake. However, it should not be assumed that logistical performance alone will always exert a lasting power of attraction on e-consumers leading them to remain loyal to a website. In other words, are we really sure that the logistical performance associated with a website creates a source of competitive advantage?

This paper wishes to underline three paradoxes tending to put in perspective the importance e-grocers should give to logistics in their business plans. As a reminder, the word paradox comes from the ancient Greek *paradoxos*, meaning “contrary to the general opinion.” Following early academic research on e-grocery

logistics, a consensus—a kind of general opinion—seems to arise on several points. We wish to adopt a critical attitude toward this consensus. Each of the paradoxes offers food for thought from elements drawn from literature, and ends with a research question making up one of the components of a research agenda to develop on e-grocery logistics. The objective is not to provide final answers on its future, but rather to state how it is possible to moderate the influence of logistical performance on the food e-tailing industry's efficiency. To do so, the paper is divided into three parts successively summing up the three identified paradoxes, i.e. that the general opinion on logistics as a key factor of success for e-grocers is questioned:

- (1) By the forced sharing of logistical resources, under the pressure of local Authorities wishing to avoid the anarchic development of private “last mile” organizations;
- (2) By the existence of a package deal of services underlining the major importance of other elements in the offer of e-grocers (privacy, sales promotion, security, etc.); and
- (3) By a level of e-consumers' expectations in terms of service quality not necessarily requiring “order fulfillment and order delivery excellence” on the part of e-grocers.

The research program whose basis is suggested here starts with the idea that several research subjects on

e-grocery logistics developed since the end of the 1990s, each leading to independent interesting and operational results. Among other things, it seems undeniable that decision-makers can now rely on computer simulations including a set of parameters to create efficient home delivery systems. Similarly, in matters of urban development, studies were started in France, in Germany and in the UK, among others, to measure the environmental impact that a rapid growth of food e-retailing could have [3] [6], and to consider a better planning of flows through the intervention of local Authorities (coordinated implementation of warehousing and transport infrastructures). If isolated information is therefore available on trends in e-grocery logistics, a comprehensive approach based on the integration of the various elements produced by a pluridisciplinary research program is still to appear. The present paper is a first step in this direction.

2. THE FIRST PARADOX—A FORCED SHARING OF THE LOGISTICAL ADVANTAGE

E-grocery (or food e-tailing) obviously questions traditional logistics, at least the final stage of (physical) contact between customers and the product assortment: it forces firms to solve the thorny problem of conveying products to e-consumers' homes —or places of work— in satisfactory conditions of cost and quality of service. Works on e-grocery logistics in the last five years have made it possible to identify the different available alternatives at order fulfillment and order delivery level [6] [10] [26]. Concerning order fulfillment, e-grocers choose warehouse picking, in-store picking or hybrid

in-store/warehouse picking. As for order delivery, attended delivery, unattended delivery and consumer pickup in third party locations all occur. No model seems to emerge durably for the time being in European countries [8]. Each food e-tailer operates with its own organization. But it is important to note that most firms offer attended deliveries based on warehouse picking, which enables e-consumers to choose a time window to receive their products ordered online (see Table I).

All observers of the e-tailing industry consider the organization of home delivery of groceries complex and expensive, whatever the order fulfillment model chosen: to contact customers, it is necessary to organize delivery rounds to service residential areas that are often uneasy to access, in limited time windows (evening, beginning of week-end). Of course, several alternatives to conveying goods to e-consumers' homes or places of work are conceivable, for example the implementation of collection and delivery points (CDP) or the use of convenience stores where e-consumers would recover the ordered products. Browne [1] gives several concrete illustrations for small packages: service stations, leisure centers (like swimming-pools or tennis clubs), railway or bus stations, etc. But is it likely that e-consumers would systematically accept using CDP or convenience stores when using e-grocery shopping is precisely a way of getting rid of such time-consuming logistical tasks? In fact, the generalization of home delivery remains the most likely scenario for the evolution of e-grocery logistics [9]. Hence the attention given to order fulfillment decisions in the academic literature (for a summary, see [26]).

Table I
Alternative e-grocery logistics models for a few firms
(adapted from Hays et al. [10])

	Warehouse picking	In-store picking	Attended delivery	Unattended delivery	In-store pickup
<i>Alberstons</i>	█	█	█		█
<i>Cold Storage</i>	█		█		
<i>Fairprice</i>	█		█		
<i>Grocery Gateway</i>	█		█		
<i>Homegrocer</i>	█		█		
<i>NetGrocer</i>	█			█	
<i>Peapod</i>	█	█	█	█	
<i>Safeway/GroceryWorks</i>	█		█		
<i>Sainsbury's</i>	█		█		█
<i>Streamline</i>	█		█	█	
<i>Tesco</i>		█	█		

The main limitation of works on e-fulfillment decisions is that they stay within a very technical view of the problem, practically in a perspective of operational research. An example would be the work of Wittgreffe & Memmory [29] who review the information

technologies best adapted to optimize the store range in the in-store picking model, or also de Koster's [17] paper on distribution structures for food home shopping. To suggest the best solution to food e-tailers (warehouse picking vs. in-store picking), de Koster [17] introduces

the notion of operational complexity, measured through the number of weekly orders, the product assortment width and the product assortment type (presence or absence of frozen and fresh products, in addition to non-perishable products). Although these studies present an obvious interest at micro-economic level, they ignore –as many others do– the macro-economic implications of decisions made by the different firms. But it is impossible to ignore a global vision of the e-grocery logistics stakes. Without such vision, there is a high risk of development of overcapacities in logistical equipment, each e-grocer investing in its own distribution network. The overcapacities will have a social cost which will quickly become unbearable in the eyes of e-consumers who are also citizens. For example it would lead to congestion of residential areas with an anarchical multiplication of delivery trucks, leading to environmental (CO₂ emissions, recurrent traffic jams) and economic problems (inability to meet time windows).

At the moment, food e-tailing represents only a low percentage of consumer sales in Europe. For example, in France, the turnover of the four major e-grocers equals the average turnover of a single large hypermarket! But national and local Authorities prefer reflecting on the potential impacts of a rapid and uncontrolled development of sales of groceries by Internet. There seems to be a desire to control product flows in towns and cities by obliging firms to use shared logistical infrastructures, at warehousing and transport level: “*Urban land use planning policy can control the number and location of home delivery fulfillment facilities... and the times at which home delivery vehicles can operate at them. Planners can also decide whether there is a role for the urban authority in the development and operation of such facilities, and whether they will be operated by one or many companies*” [2, p.34]. This planning, imposed for environmental and ecological reasons, could therefore unify the level of logistical performance, since all firms would have to use the same integrated freight centers at the same time in the same urban area [11]. If the Authorities’ interventionism in e-fulfillment is confirmed, with food e-tailers unable to co-ordinate and avoid the multiplication of private “last mile” organizations, it will be obviously difficult for a given food e-tailer to use his potential logistical excellence to distance his competitors. This is the first paradox of e-grocery logistics: it is presented as a source of competitive advantage for a firm, but the sharing of resources is going to spread this competitive advantage to all firms.

Research question No. 1: Taking into account the increasing involvement of local Authorities in the organization of urban logistics, what elements in the logistical organization will enable food e-tailers to create a difference and develop a competitive advantage?

3. THE SECOND PARADOX–THE REAL IMPORTANCE OF LOGISTICS IN THE E-MARKETING MIX

As Ricker & Kalakota [27] stated, e-consumers do not buy a product, but a product plus the “service envelope” accompanying it. Logistics is obviously one of those services, but not the only one. Logistics is but one component of a package deal, the quality and the coherence of which are crucial elements that will condition the positive or negative view of a website by e-consumers. Through the package deal of services, e-consumers try to find a turnkey solution to their problems, for example limited opportunities for in-store shopping because of a busy work schedule or occasional or lasting health problems making it difficult to regularly handle heavy or bulky products [21]. Here again we find the bases of a strategic service vision developed more than twenty years ago in academic literature [19], which retains all its relevance in a computer-mediated environment.

To state that logistics is one element of a package deal of services does not seem very original, at least when referring to works conducted on services marketing. They note that consumers acquire not a product or a service, but a more or less complex “goods-service:” any service relies on material elements which make service construction and delivery possible, any product is finally assessed in function of the service rendered. This is why it seems pertinent to reason from the generic concept of services associated with products: “*Services associated with products are services supplied in complement of a product so as to optimize their use and increase their value for customers... Expected by customers, they induce the demand for products and are the source of differentiation in firms’ offers*” [9, p.99]. The approach was applied to the French case by Paché & Bonet [23], among others, to find out whether food e-tailers try to develop the various elements of the package deal of services harmoniously or are on the contrary focusing on some of the elements (if so, why).

The interest of Furrer’s [9] analysis lies in clearly dividing the services associated with products into two categories. The first category is made of *value-added services*, relatively independent from the products they accompany and representing a sort of bonus (for example, the possibility for consumers to use a call center 24 hours a day seven days of the week). The second category is made of *services acting as barriers to entry*, an integral part of the basic offer and which must absolutely be offered under penalty of losing customers (for example, free parking space offered by large retailers). This approach is completely applicable to online purchases on the Internet. In addition to a simple product as tangible goods, e-consumers will surely investigate the overall quality of the complex service offered to them, associating after-sales support and warranties. According to Kotzab & Madlberger [18],

the package deal of services refers to three complementary dimensions –or functions– that firms should attempt to develop equally:

- A product assortment function;
- A logistical function; and
- An advice, communication and financing function.

There is no doubt that logistical performance is one of the services associated with products, but e-grocers must not neglect other at least as critical components, such as payment terms and conditions, the number of product lines offered or online advice. In a way, this echoes the notion of the e-marketing mix proposed by Kalyanam & McIntyre [14], summed up by the acronym $4Ps + P^2C^2S^3$, where 4Ps stands for the traditional marketing mix, P^2 stands for personalization and privacy, C^2 stands for customer service and community, and S^3 stands for site, security and sales promotion. Although Kotzab & Madlberger [18] do not use the notion of e-marketing mix, this is what they analyze. It is evident from this that a strict process of marketing audit, should make it possible to put in perspective the importance of services that the trade press often presents as being value-added (security, customer service, etc.), when they perhaps only serve as barriers to entry. It should be noted that some current academic studies ask a rather similar question: to what extent does an efficient logistical service contribute to the loyalty of e-consumers to e-grocers? For the time being, no clear answer has been given. A qualitative study conducted with the key informants of ten retailers in a major metropolitan area located in the USA shows for instance that they do not really know much about e-consumers'

needs and wants in terms of Internet shopping [5].

In the end, food e-tailers need to imagine a balanced e-marketing mix –whose variables reinforce themselves mutually– rather than attach themselves to the sole best logistical service possible. It is quite clear that e-consumers look for the performance of the online shopping transaction itself [25], a performance that will come from a set of closely interactive elements. The food e-tailers' objective will be to offer the highest delivered value, i.e. the most important differential between the total benefits to consumers, some examples of which are presented in Table II, and the total costs of the e-marketing offer [21]. For example, the promotional promise of a home delivery within a narrow time window to avoid long waiting times at home will require an order fulfillment and an order delivery capable of fulfilling this promise. If this is not possible, it would be better to modify the content of the promotional promise: a well-known rule of logistics management, sometimes forgotten in the computer-mediated environment. This is the second paradox of e-grocery logistics: it is most often considered as the food e-tailers' main key factor of success, whereas it is only one of the sometimes secondary components of their package deal of services.

Research question No. 2: What weight should food e-tailers give to logistical performance in their e-marketing mix, and with what other strategic variables is the place variable in direct interaction vs. indirect interaction?

Table II

*Some potential benefits to consumers offer by the e-tailers
(adapted from Jones et al. [12])*

- Convenience of comparison shopping from, and delivery to, the customer's home
- Creation of a pleasurable e-shopping experience
- Good quality and topical information on products and services
- Availability of products or services that may not be available from local traditional stores
- Potentially range of high quality products and services at competitive prices
- Speedy and reliable home delivery systems
- Online support for after-sales service
- Guarantees about the security of online transactions and the privacy of personal information

4. THE THIRD PARADOX–E-CONSUMERS' LEVEL OF EXPECTATIONS IN LOGISTICAL PERFORMANCE

An attentive reading of recent works on e-grocery shows that in fact the real influence of logistical performance in the process of e-consumers' decision-making, particularly regarding timely delivery, could be revised downward, or at least be considered as one key factor of success among others: the functionality of computer interfaces or the absolute security of payments, for example [7][22]. Cases's [3]

research on the perceived risk associated with purchases on the Internet places delivery risks far behind private risks (potential violation of e-consumers' privacy) of performance risks (products and/or services not corresponding to e-consumers' expectations). Of course, at first, when the market is gradually structuring itself, good logistics appears to be a source of competitive advantage. Will that still be the case with the rapid alignment of each website to the same cost/service standard? The difference between e-grocers will then develop on their capacity to establish, and then to maintain an individual, long-term and customized

relationship with consumers, while the traditional bricks-and-mortar model lies on a rather intermittent, distant and impersonal relationship [18].

The first field investigations conducted on e-consumers' practices underline the central nature of convenience relations that occur between for example a delivery man and an e-consumer when orders are regular. Kessous [14] lists numerous, often amusing cases of complicity building up between both agents, as in the case of a delivery man who knows that Mr. Smith, whose habits he is perfectly aware of, must receive delivery at 9 am just after his morning walk with his dog! We are here faced with the construction of a real joint service and delivery, the efficiency of which directly determines the level of satisfaction perceived by the customer. Beyond simply adhering to delivery times, the ability to customize service becomes a key factor of differentiation for e-grocers. More generally, e-consumers consider that it is normal to be delivered on time without error in order picking. But they will expect increasingly sophisticated and customized services associated with products, such as technical information, as the preliminary results of the Swedish program *Households in Cyberspace* show [26].

The central question is to know what is the level of service quality perceived as necessary and sufficient by e-consumers (in their online shopping transaction), when numerous European food e-tailers essentially try to offer better order fulfillment and order delivery than those of their major competitors. It is highly likely that e-consumers represent a heterogeneous population in their expectations about logistical performance. For some of them, the strict adherence to the time window stated on the website is capital, for others, with more available time at home during the day, the lowest possible delivery charges will make the difference. Systematically proposing high service quality standards may well be unnecessarily expensive for food e-tailers, since some e-consumers will be impervious to them. It would be better to study the expected levels of service quality per market segment beforehand, offer a highly differentiated logistical performance (in terms of delivery charges, in extended time windows, etc.), and above all to comply with what was promised. This approach puts customer concern at the heart of the analysis, focusing on operational issues that are currently largely underestimated [13]. In fact, some of the components of logistical performance may not be deemed essential by some e-consumers, and food e-tailers would make a serious management mistake if they went on considering them as a key factor of success.

This last point seems all the more important for the future of e-grocery logistics since works conducted in the USA by Morganosky & Cude [21] [22], from a longitudinal study on e-consumer behavior, show that e-consumers are above all loyal to a website, to which

they gradually dedicate an increasing part of their online shopping, playing on a phenomenon of experience. As has been said, it is the overall online transaction that provides a value of variable importance, not just the quality of the order fulfillment and order delivery processes. E-consumers gradually "learn," get to know the website's logistical performance and increase or reduce their level of expectations accordingly. Thus, when e-consumers have the possibility to control the shipping function, for example by freely choosing a particular carrier on the website, they link it to a level of logistical performance related to its image, particularly in terms of service customization via the delivery man, and adapt their delivery-related expectations accordingly [7]. In brief, an e-consumer loyal to a website for extra-logistical reasons could very well make do with a very indifferent level of service quality, compared to other websites. This is the third paradox of e-grocery logistics: the total efficiency of the order fulfillment and order delivery processes does not necessarily have a significant impact on the satisfaction felt by e-consumers.

Research question No. 3: How does the level of expectations of market segments develop in terms of logistical performance, and to what extent does a high (low) level of logistical performance influence loyalty (disloyalty) to a website?

5. CONCLUSION

In the last few years, works on e-grocery logistics have multiplied, particularly in Northern European countries, widely involved in online Internet sales. Reflections are already quite advanced on possible alternatives in order fulfillment and order delivery; they have led to extended simulations to define the best solution according to variables such as population density in the residential area to serve, the frequency and size of orders in households, or the last mile management cost [15] [26]. Without minimizing the interest of this approach, it should be admitted that it relies on an implicit but debatable postulate: e-consumers' high awareness of logistical performance and, by extension, the crucial nature of "logistical excellence" for e-grocers wishing to get ahead of their competitors in the long-term. The present paper questions this postulate, referring to the three paradoxes of e-grocery logistics.

The suggested research program still has to be developed in two directions. One belongs to the BtoB type: it studies the impact of the legal environment on logistical strategies. The second direction belongs to the BtoC type studying the place of e-grocers' logistical performance in e-consumer behavior, compared to the place held by other variables in the e-marketing mix. This will become a reality only if it is possible to eliminate a number of disciplinary compartments, as the research agenda involves resorting to marketing tools, logistics management tools, political science and strategic management tools. Hence the importance of

building managerial teams with complementary competencies, the only way to go beyond the technical approach to e-grocery logistics still dominant in today's literature, and far too simplistic for understanding the true challenges being faced.

REFERENCES

- [1] Browne, M., "E-commerce, freight distribution and the truck industry," 4th ACEA SAG Meeting Discussion Paper, Brussels, October 2000.
- [2] Browne, M., Allen, J., Anderson, S., Jackson, M., *@ Your home: new markets for customer service and delivery*, Foresight Report, London: UK Retail Logistics Task Force, October 2001.
- [3] Cairns, S., "Home delivery: the environmental consequences," TSU Working Paper No. 1999/5, London: University College London, March 1999.
- [4] Cases, A.-S., "Perceived risk and risk-reduction strategies in Internet shopping," *International Review of Retail, Distribution and Consumer Research*, Vol. 12, No. 4, pp375-394, 2002.
- [5] Cowles, D., Kiecker, P., Little, M., "Using key informant insights as a foundation for e-retailing theory development," *Journal of Business Research*, Vol. 55, No. 8, pp629-636, 2002.
- [6] Durand, B., *La logistique de l'épicerie électronique: analyse prospective des scénarios envisageables dans la distribution alimentaire française*, Thèse de doctorat en Sciences de Gestion, Université de la Méditerranée (Aix-Marseille II), December 2003.
- [7] Esper, T., Jensen, T., Turnipseed, F., Burton, S., "The last mile: an examination of effects of online retail delivery strategies on consumers," *Journal of Business Logistics*, Vol. 24, No. 2, pp177-203, 2003.
- [8] Fernie, J., Fernie, S., Moore, C., *Principles of retailing*, Oxford: Butterworth-Heinemann, 2003.
- [9] Furrer, O., "Le rôle stratégique des 'services autour des produits'," *Revue Française de Gestion*, No. 113, pp98-108, 1997.
- [10] Hays, T., Keskinocak, P., Malcome de López, V., "Strategies and challenges of Internet grocery retailing logistics," in Geunes, J., Akçali, E., Pardalos, P., Romeijn, H., Shen, Z. (eds.), *Applications of supply chain management and e-commerce research in industry*, Dordrecht: Kluwer Academic Publishers, pp217-252, 2004.
- [11] Hesse, M., "Land for logistics: locational dynamics, real estate markets and political regulation of regional distribution complexes," *Tijdschrift voor Economische en Sociale Geografie*, Vol. 95, No. 2, pp162-173, 2004.
- [12] Jones, P., Clarke-Hill, C., Hillier, D., "(R)etailing in the UK," *Marketing Intelligence & Planning*, Vol. 24, No. 4, pp229-233, 2002.
- [13] Jones, P., Clarke-Hill, C., Shears, P., Hillier, D., "The eighth 'C' of (r)etailing: customer concern," *Management Research News*, Vol. 24, No. 5, pp11-16, 2001.
- [14] Kalyanam, K., McIntyre, S., "The e-marketing mix: a contribution of the e-tailing wars," *Journal of the Academy of Marketing Science*, Vol. 30, No. 4, pp487-499, 2002.
- [15] Kämäräinen, V., *The impact of investments on e-grocery logistics operations*, Dissertation for the degree of Doctor of Science in Technology, Helsinki University of Technology, June 2003.
- [16] Kessous, E., "Le commerce électronique et la continuité de la chaîne logistique: de l'approvisionnement des sites à la livraison aux consommateurs," *Réseaux*, Vol. 19, No. 106, pp104-133, 2001.
- [17] de Koster, R., "Distribution structures for food home shopping," *International Journal of Physical Distribution & Logistics Management*, Vol. 32, No. 5, pp362-380, 2002.
- [18] Kotzab, H., Madlberger, M., "European retailing in e-transition? An empirical evaluation of Web-based retailing—indications from Austria," *International Journal of Physical Distribution & Logistics Management*, Vol. 31, No. 6, pp440-462, 2001.
- [19] Levitt, T., "Marketing success through differentiation—of anything," *Harvard Business Review*, Vol. 59, No. 3, pp83-91, 1980.
- [20] Licoppe, C., "Pratiques et trajectoires de la grande distribution dans le commerce sur Internet: vers un autre modèle de coordination pour le commerce électronique?," *Revue Economique*, Vol. 52, HS, pp191-211, 2001.
- [21] Morganosky, M., Cude, B., "Consumer demand for online food retailing," *Journal of Food Distribution Research*, Vol. 32, No. 1, pp5-17, 2001.
- [22] Morganosky, M., Cude, B., "Consumer responses to online food retailing: is it really a supply side issue?," *International Journal of Retail & Distribution Management*, Vol. 30, No. 10, pp451-458, 2002.
- [23] Paché, G., Bonet, D., "Le service logistique est-il perçu comme créateur de valeur par les entreprises du commerce électronique? L'expérience des distributeurs alimentaires en ligne français," *Proceedings of the 2003 AMA ServSig Services Research Conference*, Reims, pp1-15 (CD-rom), 2003.
- [24] Pandya, K., Arenyeka-Diamond, A., "Towards a model of e-tailing: a SWOT analysis of electronic supermarkets," *International Journal of Services Technology and Management*, Vol. 3, No. 1, pp96-110, 2002.

- [25] Park, C.-H., Kim, Y.-G., "Identifying key factors affecting consumer purchase behavior in an online shopping context," *International Journal of Retail & Distribution Management*, Vol. 31, No. 1, pp16-29, 2003.
- [26] Punakivi, M., *Comparing alternative home delivery models for e-grocery business*, Dissertation for the degree of Doctor of Science in Technology, Helsinki University of Technology, June 2003.
- [27] Ricker, F., Kalakota, R., "Order fulfillment: the hidden key to e-commerce success," *Supply Chain Management Review*, Vol. 3, No. 3, pp60-70, 1999.
- [28] Wikström, S., Persson, C., "Home shopping: once the novelty wears off," *ECR Journal*, Vol. 1, No. 1, pp27-35, 2001.
- [29] Wittgreffe, J., Memmory, R., "Iceland Internet shopping," *BT Technology Journal*, Vol. 18, No. 2, pp72-78, 2000.

Research on Equipment and Spare Parts Management Based on Theory of Constraints

Yan Ye

Doctoral candidate, Department of Industrial Engineering and Management, Shanghai Jiao Tong University,
Shanghai 200030, China

Institute of Industrial Engineering, Zhejiang University of Technology, Hangzhou 310032, China
yeyna@sjtu.edu.cn

ABSTRACT

On the basis of analyzing the relationship between the equipment system and spare parts inventory, this paper proposes two-level equipment and spare parts system management based on the Theory of Constraints. From the system point of view and according to the Theory of Constraints, the manufacturing organization identifies bottleneck equipments in equipment system and makes the bottleneck equipments operate continuously and effectively on the basis of two important aspects of materials supply and spare parts supply. Finally, the spare parts' ABCD-TOC inventory classification and control method based on the Theory of Constraints is introduced. This method classifies spare parts according to different demands of bottleneck and non-bottleneck equipments, technical rules of spare parts failure and waste and purchase characteristics. Then the suitable control strategies for each class are adopted to focus on bottleneck equipments without sacrificing the spare parts supply for non-bottleneck equipments, as well as to reduce inventory investment.

Key words: Theory of constraints, Equipment management, Spare parts management, Inventory

1. INTRODCUTION

During the past two decades, as a result of serious global competition, the modern company depends on the high-effective equipments increasingly to strengthen its competitive capacity. And along with the development and progress of science and technology, equipments tend to the large-scale, high precision and complexity and their automation degree has enhanced. The maintenance expenses and fault loss of equipments account for an increasing fraction of the production costs. It has a great effect on the competitive ability that companies keep and improve equipment utilization. But equipment utilization is decided not only by its reliability and maintainability but also by spare parts supply which is the essential physical foundation^[1]. The spare parts management also is an important part in management activities of company equipment.

Spare parts inventories differ from other manufacturing inventories in several ways. First, the functions are different. Raw material inventories exist to offer the required raw materials for production. Work-in-process (WIP) inventories exist in order to smooth out irregularities in production flow. Finished product inventories exist to provide a source of products for delivery to customers and are designed to protect against irregularities in differences in quality levels, scheduling problems, and other well-known production characteristics. The function of spare parts inventories, however, is to assist a maintenance staff in keeping equipment in operating condition and promoting continuous production process. Second, WIP and final product inventories can be increased or decreased by changing production rates and schedules, improving

quality, etc. Spare parts inventory levels, however, are largely a function of how equipment is used and how it is maintained. Therefore the spare parts inventory and the status of equipments promote mutually^[2].

For avoiding excessively long equipment downtime due to lack of spare parts required and corresponding loss of time, output and profit, companies always prepare adequate spare parts to promptly supply. However, inventory is expensive especially when spare parts are stocked too much and the total investment in spare parts is very large. Therefore there is the conflict of spare parts inventories, equipment operation and investments for the goals of enhancing competitive ability and realizing high profitability (see Fig.1). Managing and controlling effectively equipment and spare parts system may be of vital significance to balance these goals of maintaining high equipment availability, minimizing inventory investment and attaining high profitability.

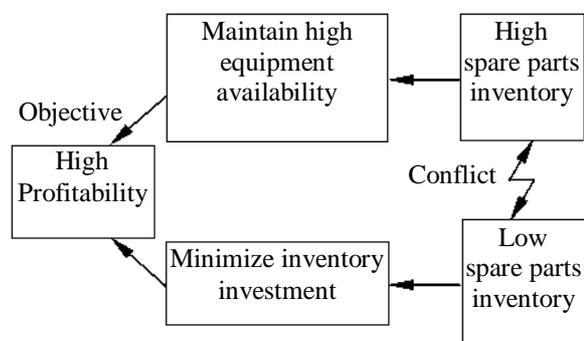


Fig. 1. Relationship between spare parts inventory and equipment system

The Theory of Constraints (TOC) on a basis of

Optimized Production Technology (OPT) was expounded by Israeli physicist Eliyahu M. Goldratt (1986), which was developed from initial manufacturing management principle to "Thinking Process" of solving problems logically and systematically. The theory was applied to many fields beyond production management [3].

The organization of this paper is as follows. The basic concept of TOC is sketched in Section 2. Next, Section 3 proposes two-level equipment and spare parts system management based on the Theory of Constraints. On a basis of analyzing two important supporting resources for equipment operation, the spare parts' ABCD-TOC inventory classification and control method is introduced. A summary and concluding remark is given in the final section.

2. BASIC PRINCIPLE OF TOC

The Theory of Constraints considers the production and operations of manufacturing organizations from the entire system point of view. Goldratt argues that the goal of a company is to make money and develops a different set of measurements which express the goal: throughput, inventory and operational expense. They are defined as follows. Throughput is the rate at which the system generates money through sales. Inventory is all the money that the system has invested in purchasing things which it intends to sell. Operational expense is all the money the system spends in order to turn inventory into throughput [4].

In any system, if any link output is decided by the outputs of one or several stages before it, the link with the least output determines the output level of the whole system, that is, the strength of a chain is determined by the weakest link. The link is called a constraint and also a bottleneck. TOC is the management idea and principle that help companies to identify, exploit and eliminate constraints and then to maximize throughput and reduce inventory and operational expense. Companies can apply the Theory of Constraints through five-step focusing process as follows [4]: (1) Identify the system's constraint(s); (2) Decide how to exploit the system's constraint(s); (3) Subordinate everything else to the above decision; (4) Elevate the system's constraint(s); (5) If in the previous steps a constraint has been broken, go back to step 1, but do not allow inertia to cause a system's constraint.

The process of implementing these five steps is also three steps thinking process: "what to change?", "what to change to?" and "how to cause the change?". Then, companies can achieve their goal by focusing on the key point-the constraint and can promote continuous improvement because the constraint is not static but dynamic and changing.

3. TWO-LEVEL EQUIPMENT AND SPARE PARTS SYSTEM MANAGEMENT BASED ON TOC

3.1 General Framework

Market competition is more and more fierce and customers' demands for diversified products tend to enhance, which cause shorter product life cycle and decrease product quantities of each order [5]. For abstaining from risks, companies do not purchase or update expensive production resources easily and exploit the utilization of existing equipments as far as possible. So the overall productivity appears insufficient in the growing and mature period of product life cycle. According to TOC, we regard several or a set of equipments producing one kind of product or a series of products as an equipment system. In this system if any segment (the equipment or equipment group) hinders the overall system to increase throughput or decrease inventory and operational expense, the segment is a constraint, that is, the bottleneck equipment. The bottleneck equipment whose capacities are not sufficient restricts the system throughput and decides the degree of other non-bottleneck resources utilization. As the physical base of equipment system, the main task of spare parts inventory is to assure all equipments in good condition. A spare part, especially for bottleneck equipments, is called the bottleneck spare part so long as it affects the system throughput, which is the important point of the spare parts management. The structure of two-level TOC-based equipment and spare parts system management is depicted in Fig. 2.

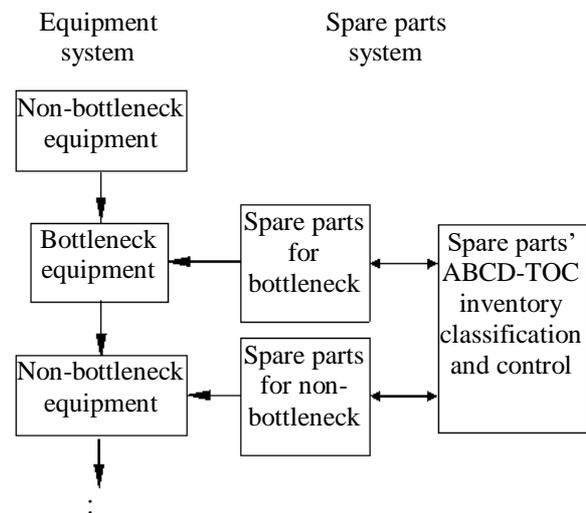


Fig. 2. Two-level equipment and spare parts system management based on TOC

3.2 TOC-based equipment system management

The fundamental task of TOC-based equipment system management is to identify the bottleneck of the entire system and manage the bottleneck and resources before

and after it to maximize system throughput. The bottleneck equipment of the equipment system is the equipment or equipment group whose actual capacities are smaller than or equal to production loads (demands), that is, the system's constraint is the equipment experiencing the highest load rate. The equipment load rate is defined as its duty to its available capacity during the plan period, which is given by equation (1).

$$L_t = \frac{\sum Q_t S}{TA_t} \times 100\% \quad (1)$$

Where L_t is the equipment load rate over the plan period t . Q_t represents quantities of each product that the equipment should produce during t . S is defined as the planned time standard of each produce. T indicates the institutional time of the equipment during t . A_t is the probability of equipment availability which comprehensively reflects reliability and maintainability.

After the bottleneck equipment is identified, how to make full use of the equipment is the first problem that needs to be solved. Considering the direct resources required to produce products, we may establish the material buffer to make equipment run continuously. The buffer is often classified into inventory buffer and time buffer. Inventory buffer is a kind of insurance Work-in-processes. Time buffer is required materials handed over advancing a period of time from the planned time to eliminate influences of stochastic fluctuation in production processing, which is measured by processing time length of bottleneck. On the other hand, from the technical resources supporting equipment system viewpoint, we should increase the overall equipment effectiveness (OEE) and reduce unscheduled downtime especially caused by equipment failure as far as possible, which depends on maintenance resources such as maintenance men, maintenance tools and supply of spare parts. However, compared with the service men and tools, spare parts investment is higher and lack of spare parts is the most main influence factor frequently. Thus, this paper introduces TOC-based spare parts inventory management to keep equipment, especially bottleneck equipments, run well and maintenance efficiency as well as reduce costs.

3.3 TOC-based spare parts inventory management

Generally companies have large quantities of and many types or specifications of spare parts inventories. To manage inventories through Stocktaking, placing and accepting orders endlessly would be time-consuming. Then companies manage spare parts inventories through classification.

Dickie proposed ABC classification method ^[6] in 1951. The purpose of this method is to measure and classify the degree of importance and concentration of inventory

items in terms of the value of items and Pareto principle. The method is simple and wieldy and so is adopted by many companies for a long time. However, the value of spare parts has no inevitable relations with their influences on the equipments and production.

In accordance with the special characteristics of spare parts inventory, 3A inventory control method ^[7] classifies spare parts into 27 different kinds of rank from AAA to CCC according to the important degree of spare parts to equipments. Considering actual operations conditions of equipments, the method decides essentiality and specific rank and order of each spare part according to three aspects: equipment and its components belonging to, production load enduring and its effect on product quality. Then the method adopts the corresponding control method. Guo Zhimin et al. ^[7] offer ABCD inventory control method in terms of several criterions such as influence of spare parts on productive condition, the rules of fault and wasting, and agility of supply channels. Dekker et al. ^[8] classify demand for spare parts into critical and non-critical demands according to the influence caused by the disabled equipment that a spare part is installed. In this paper we analyze classification and control methods of spare parts from the system point of view, not only from single equipment viewpoint, and according to the Theory of Constraints that to our knowledge have not been discussed in the literature so far. The method is called spare parts' ABCD-TOC inventory classification and control method which classified spare parts into four classes.

A class includes special spare parts that are installed to the bottleneck equipments and can not be checked in advance or their failure and wear time can not be forecast, and that are installed to the non-bottleneck equipments and supplied by few providers and/or with long purchase lead time (such as purchased from abroad). Because an hour lost at the disabled bottleneck equipment is an hour out of the entire system and part of these spare parts are only used by the bottleneck equipment, this type is very important. Compared to the bottleneck equipment, non-bottleneck equipment always has idle time. But if purchasing time exceeds idle time, non-bottleneck equipment will affect production system throughput. Thus, this class may be called the bottleneck spare parts and is controlled by an $(S-I,S)$ policy, or equivalently, by a base stock policy ^[2]. The base stock S is defined as maximal possible consumption during an order lead time. Concretely, it can be calculated through adding historical average consumption over an order lead time to a suitable safety quantity. The safety quantity is achieved by multiplying fluctuation average of historical consumption by certain safe factor as a kind of buffer which is considered comprehensively according to production task, usage years of equipments, etc. Stock level will be inspected and checked after spare parts are used or continuously. When the actual valid inventory level is lower than S ,

one purchase order should be placed with the order quantities which is decided by subtracting the current stock level from S .

B class includes spare parts with predictable replacement time used by bottleneck and/or non-bottleneck equipments. Companies can forecast the time when need to replace this type spare parts through checking in advance, tracking and analyzing their conditions in time. Or they track daily usage of these spare parts in a long time and statistically analyze the rule of their fault and waste to estimate their replacement time cycle. When companies have predicted the replacement need during equipment operating period, they should estimate the earliest time and advance several days according to specific situations of the time and then order ahead of purchase lead time. Thus companies can supply required spare parts timely while reduce occupied time and fund of inventory.

C class is the special spare parts with many supply channels and/or short purchase lead time which can be sucked by idle time of equipments. For avoiding depletion in stock and reducing investment, this type can be controlled by zero inventory policy, that is, these spare parts are not stocked. When production department demands the spare parts, it applies for the purchase immediately. Then purchase department orders immediately from suppliers. After spare parts arrive, purchase department checks and accepts and warehouse registers. The spare parts are sent to the production center directly.

All general spare parts fall into D type. The general spare parts are installed at some even all equipments, so their reflections of production and maintenance change are not obvious. When a general spare part is requested by bottleneck and non-bottleneck equipments synchronously, it must meet bottleneck first that helps to assure and increase system throughput. This type has many kinds of spare parts whose usages and investments are large. The class can be controlled by periodic review (t, s, S) inventory policy^[9] where t presents fixed review period, s reorder point, S maximum stock level. We estimate s by computing consumption of each spare part during purchase lead time after examining its daily average usage over a long period of time, and take S as three or four times of s . Suitable t can be set according to the spare part with the fastest consumption rate in this type as long as all spare parts of the type are not short during the period t . Inventory levels of these spare parts are checked and compared to s every other t . If inventory level is lower than s , the spare part would be ordered and order quantity is the balance of S and current level. If the level is larger than s , there is no any order. This type of periodic review reduces massive work as a result of continuous review and keeps proper inventory level without short of spare parts and with lower investment.

It is worth explaining that the classification of A, B, C and D classes is not always changeless but dynamically adjusted and transformed one another. For example, after some measures for improving equipments capacity are taken constantly, some bottleneck equipments change to non-bottlenecks and then some spare parts of A class are transformed into C class. When more advanced statistical analysis technology make the fault and waste rule of equipment and spare parts learned, some A class spare parts may be classified into B class.

4. CONCLUSION

This paper analyzes the relationship between the equipment system and spare parts inventory and proposes two-level equipment and spare parts system management based on the Theory of Constraints. From the system point of view and according to the Theory of Constraints, the company identifies bottleneck equipments in equipment system and makes the bottleneck equipments operate continuously and effectively. Companies can take measures such as buffers to supply enough materials. However, the same important thing that companies should do is to supply enough spare parts timely. So the spare parts' ABCD-TOC inventory classification and control method based on the Theory of Constraints is introduced. This method classifies the spare parts according to different demands of bottleneck and non-bottleneck equipments, technical rules of spare parts failure and wasting and purchase characteristics. Then the suitable control strategies for each class are adopted to focus on getting bottleneck equipment operate without sacrificing the spare parts supply for non-bottleneck equipments, as well as to reduce the occupied fund of inventory. This paper explores the management of equipment system and spare parts inventory from the entire system viewpoint and it provides the theory and application values for modern companies.

REFERENCES

- [1] J.W. Rusterburg, Li Peiran (translating), "Application and economic significance of system approach for spare parts management", *Equipment Management and Maintenance*, 2003(12): pp42-43(in Chinese)
- [2] W.J. Kennedy, J. Wayne Patterson, Lawrence D. Fredendall, "An overview of recent literature on spare parts inventories", *International Journal of Production Economics* 76 (2002) pp201-215
- [3] Wang Yurong, *Bottleneck management (TOC)*, Beijing: China Machine Press, 2002(in Chinese)
- [4] Eliyahu M. Goldratt, Jeff Cox, *The goal: A process of ongoing improvement*, New York: North River Press, 1992
- [5] Daniel Sipper, Robert L. Bulfin, *Production: planning, control and integration*, The McGraw-Hill Companies, Inc., 1997

-
- [6] Roger G. Schroeder, *Operations management: Decision making in the operations function (4th ed.)*, The McGraw-Hill Companies, Inc., 1993
- [7] Guo Zhimin, Yan Hongseng, Chen Shihua, Wang xuewei, "Method of spare parts' inventory control", *Computer Integration Manufacturing System-CIMS*, 2003, 9(10): pp1028-1032(in Chinese)
- [8] R. Dekker, M.J. Kleijn, P.J. de Rooij, "A Spare Parts Stocking Policy Based on Equipment Criticality", *International Journal of Production Economics* 56-57 (1998) pp69-77
- [9] Hu Yunquan, *Operations research tutorial*, Beijing: Tsinghua University Press, 1998(in Chinese)

Semantic Innovation Management

Jianqiang Li, David O'Sullivan

Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland
{Jianqiang.Li, David.osullivan}@deri.org

ABSTRACT

Innovation within industrial environment can be viewed as a cyclic loop consisting of four distinct phases, i.e., recognition, initiation, implementation, and stabilization. Different information technology enabled innovation management tools supporting the lifecycle of innovation are classified as five layers, i.e., individual innovation, project innovation, collaborative innovation, distributed innovation, and semantic innovation. According the fact that the current state is evolving from distributed innovation to semantic innovation, this paper focus on the realization of Semantic Web technologies enabled semantic innovation. To explicitly and formally specify all the different perspectives of innovation related information, a shared ontology is proposed as the common language of innovation management, which describes the critical and minimal information about the innovation process in a holistic way. Then, a technical framework which employs the machine readable innovation ontology to actually improve innovation management inside an organization and among loosely coupled organizations is presented. Finally, some features of the semantic innovation are discussed.

Keywords: Innovation management, Semantic Web, Innovation ontology, Goals collaboration

1. INTRODUCTION

Innovation [1] is “the process of making changes in something established by introducing something new”, which can be viewed as a cyclic loop consisting of four distinct phases, i.e., recognition, initiation, implementation, and stabilization. Within a typical organisation, there are various degrees of innovation from simply suggesting ideas or managing a single project to managing an entire programme of change comprising various performance indicators and portfolios of projects. To support the lifecycle of innovation, different innovation management tools or tool families have been developed. Conceptually, they can be classified as five layers, i.e., individual innovation, project innovation, collaborative innovation, distributed innovation, and semantic innovation.

Individual innovation concentrates on the development of effective ways to enhance creative potential for individual work. There are many tools or techniques available (e.g., mind mapping, brainstorming, cause-effect diagrams, Delphi forecasting) to support individuals and small groups engaged in single innovation tasks such as problem solving and idea generation. Project innovation concerns about the systematic techniques that support the change manage of project teams in a distributed asynchronous environment. This area is currently saturated with various software tools that support distributed project management and information sharing between distributed individual or small groups in a project teams. However, very few tools explicitly link project actions (e.g. tasks) with project goals. Collaborative innovation focuses on the change management among a large group of individuals within a whole organisational unit and where there is typically a portfolio of projects (rather

than one single project). In this type of innovation, the organizational unit will also typically have multiple goals such as a strategic plan, group of performance indicators or international standard, and the goals of the organization unit replace project goals. Current research focuses on the development of structured collaboration environments, and the entire system is designed to be used collaboratively by every user of the system to gather information and update their own contributions to their community. Distributed innovation is innovation across a particular intranet within an organizations supply chain and even a specific virtual team. This level of innovation is defined by all of the ‘collaborative’, ‘project’ and ‘individual’ innovations, taking place lower in the innovation hierarchy. The tools required for distributed innovation, which build mostly on portal servers (e.g., Sharepoint, Lotus Notes, and Plumtree), are in their infancy. Principal design features include enterprise wide search and navigation, user personalisation, integration and content management, notifications (push technologies), workflow, and application integration to other information sources. Since current state of the art is manual browsing, a well designed collaboration/management structure is necessary to allow any individual easy access to the information that they are looking for.

The generation of semantic innovation is derived by the emerging semantic web technology. According to Tim Berners-Lee [2], the Semantic Web can be envisioned as an extension of the current web, which makes the web more understandable to computer programs, and then allows data to be shared and reused across application, enterprise, and community boundaries easily. The original intention the proposition of semantic innovation is to exploit the potential of semantic web technology to improve the effectiveness and efficiency of innovation

management in the large networked organization.

2. INNOVATION ONTOLOGY

An ontology [3] is a general conceptualization of a specific domain in a both human and machine understandable format. In general, it consists of classes, properties, relationships, and axioms. As the backbone technology for the Semantic Web, ontology promise a share and common understanding of a domain that can be communicated between people and application.

To realize semantic innovation, it is necessary (as the prerequisite) to build a formally and explicitly expressed conceptual framework (i.e. ontology) about innovation related information. It should include all the necessary information for innovation management to facilitate the collaboration and innovation management in a networked organization and provide the conceptual underpinning for making the semantics of innovation related metadata machine interpretable.

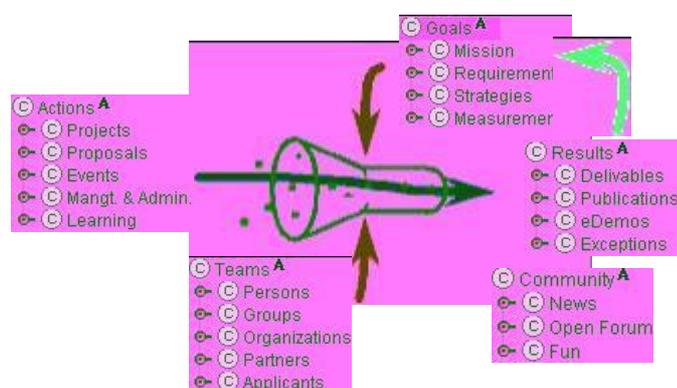


Figure 1. Conceptual Framework for Innovation Management

Considering the fact that different organizations have different concepts and relationships of their work activities and context, but on the other hand they all have a set of innovation related information for their very own, local (private) and shared (public) ontologies are distinguished here. The semantics of an organization's specific working context is captured by its local or private ontology which serves the problem solving purposes (particularly in communication and information exchanges) of this organization. A public or shared ontology are used to articulate the intersectional semantics about a common subject (or domain) which need to be shared and communicated across multiple local contexts. The main role of shared ontology is to support the common subject related interoperation across multiple organizations. Obviously, the ontology we try to build here is a shared ontology and the common subject refers to innovation management.

Since innovation can be visualised as the interaction of various types of information and in particular the organisations Goals, Actions, Teams and Results. The ontology for innovation management derived from the

iTeam methodology [4] includes five correlated concepts or classes as shown in Fig. 1.

In this innovation management oriented ontology, all the information about innovation management is classified as five categories and specified as five abstract classes, i.e., Goals, Actions, Teams, Results, Learning, and Community. Goals are defined as the objectives of a networked organization's effort. It can be embodied as Mission, Requirements, Strategies, and Measurements. Obviously, these four subclass are closely relevant each other. For example, the performance indicators of the Measurements are the quantitative estimation of the Strategies. Actions are defined as expenditure of the effort to achieve the goals, which are materialized as Calls, Proposals, Projects, Events, Learning. To realize the alignment of goals and actions, the concept of Goals should be linked with the concept of Actions. For example, a projects is aligned with which strategies and how to measure the performance of the project. Teams describe the human elements of the organization that interact within the innovation process and permit the effective management of their involvement. It includes five subclass, i.e., Persons, Groups, Organizations, Partners, and Applicants. All individuals should be linked to the goals of the development process through an inbuilt performance appraisal system. In fact, most parts of the Teams classes are inherited form FOAF and extend FOAF with innovation related information, such is each person or organization has a property representing his/its goal. Results specify the extensive yet concise reporting of outcome for all goals and actions, which are refined as for subclasses, i.e., Publication, eDemos, Deliverables, and Exceptions. The results area deploys effective management techniques such as the 'traffic lights' metaphor and control charts that allow results to float within predefined limits before they attract attention. Community provides team members with different collaboration spaces for their different interest. Currently, it is divided into News, Fun, and Open Forum, which are necessary supporting tools for goals directed collaboration.

Using a same conceptual framework to formally specify different perspectives of innovation related information extracted from different specific organization context, the innovation ontology can serve as the common language to describe the critical and minimal information about the innovation process in a holistic way. It is independent of the specific work context whichever the corresponding organization resides in and can be constructed from scratch or by composing selected parts of existing ontologies in the networked organization.

It is generally accepted that it is impossible to achieve a comprehensive and globally consistent ontology about the world [3]. In fact, it is even difficult to build a generally accepted ontology about a specific domain. The innovation ontology built here is mainly suitable for

semantic innovation of a research institute. But it can easily be adapted to the semantic innovation of other kinds of organization or other levels (e.g., 'collaborative', 'project', and 'individual' innovations) in the innovation hierarchy.

The ontology in which information is given well-defined meaning and accessible to potential semantic web applications will enable computers and people to work cooperatively and efficiently in the innovation process. It also serves as the base for the design and implementation of specific tools or application for semantic innovation management.

3. A FRAMEWORK FOR SEMANTIC INNOVATION

Creating such an ontology—a consensual understanding of the key terms used in innovation management is only the first step to realize semantic innovation management. How to employ the machine readable metadata to actually improve innovation management inside an organization and collaboration among loosely coupled organizations is the next step to be considered. A framework for exploiting the potential of semantic web technology to improve the flow of innovation process is proposed as Fig. 2.

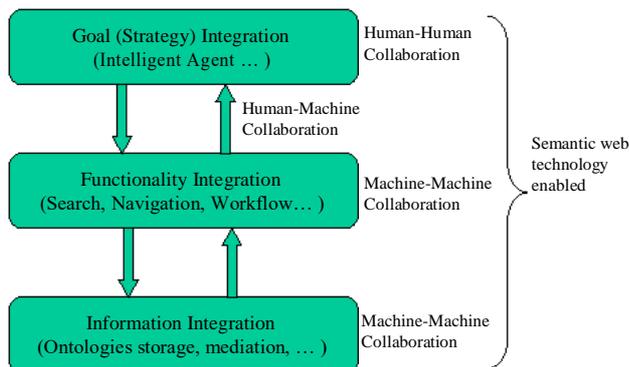


Figure 2. A General Framework for Semantic Innovation

The technical framework for semantic innovation includes three levels, i.e., information, functionality and goal (strategy) integrations. Ontology-based information integration is the foundation of the semantic innovation. Relevant information about innovation management are marked up using innovation ontology, then semantic web based application existing in the level of functionality integration can better understand the semantics and therefore more intelligently locate and integrate data for a wide variety of organizations or projects. Through ontologies storage, mediation, evolution management, it facilitates the smooth flow of relevant information among different involved parts of innovation process. To use the machine accessible information model to realize semantic innovation management needs specific application (e.g., semantic inference, context aware search and navigation, workflow management, and so on) that allow

'machine-machine'interrogation and interpretation of innovation related data. That is what function integration level cares about.

It employed the integrated semantic information to enable the different functionalities and their interoperation required by corresponding innovation process. To encourage the different levels members (individual, project, an organization) to participate in the interactive innovation process, the prerequisite is that there are shared goals or interdependent goals existing in these members. The high level goal integration is the guarantee for the realization of low level information and functionality integration. The goals alignment in distributed innovation is conducted by the human intuitively and manually. However, in the semantic innovation process, intelligent agents specified in terms of the respective goals and obligations are employed as representative of corresponding master (an individual, a project, or an organization) to cooperate with one another to realize goal directed collaboration.

Obviously, goal integration corresponds to human-human collaboration, functionality and information integrations correspond to machine-machine collaboration. The interaction between goal and functionality corresponds to human-machine collaboration. Obviously, the ultimate goal is to improve goal integration by effective and efficient machine-machine collaboration.

3.1 Interface

According to [2, 5], the integration of agent technology and ontologies could significantly affect the use of Web services and the ability to extend programs to perform tasks for users more efficiently and with less human intervention. Agent is naturally selected as the interface for the human-human and human-machine collaboration in semantic innovation process (In fact, software agent is also the generally accepted user interface of Semantic Web). The effectiveness of such software agents will increase exponentially as more machine-readable Web content and automated services (including other agents) become available. In the scenario of semantic innovation, different intelligent software agents work together in anticipating user's information requirements and thus avoid manual browsing for common information gathering tasks. The shared ontology allows for the development of search tools and intelligent agents that can automatically find any information requested by the user and thus avoid inefficient or manual 'surfing'. With semantic innovation any user will have instant access to all of the innovation going on anywhere within the organisation, regardless of language, structure, or location of the information. These semantically rich applications will be capable of real-time event monitoring and handling alerts, ensuring the stakeholder is constantly engaged and provided with priority data.

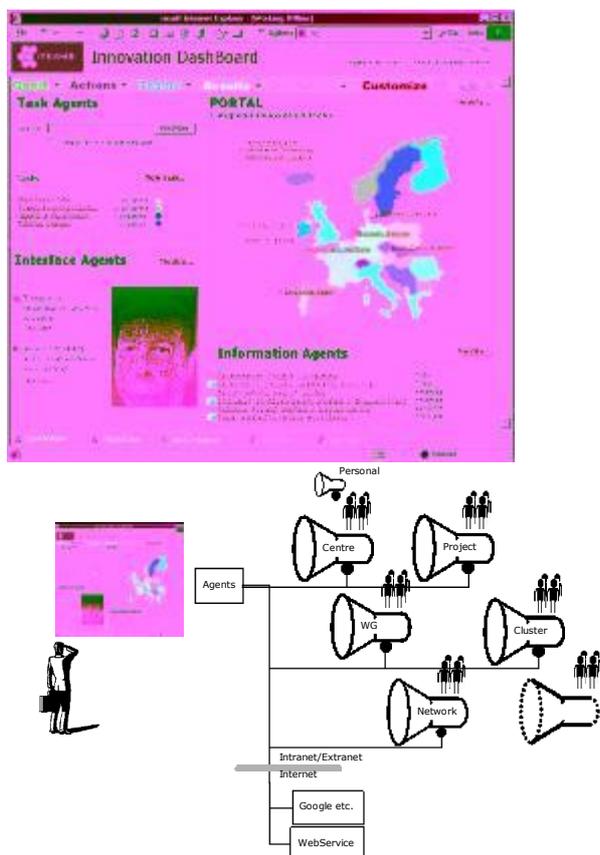


Figure 3. Interface of Semantic Innovation Management

Figure 3 presents a schematic representation of semantic innovation. The user has available to him or her a wide range of innovation funnels –some dealing with product innovation, others based on individual projects and others perhaps for the specific development plans of the corporation. In this vision a number of software agents are deployed across the corporations intranet to communicate with each other regarding the exchange and sharing of innovation information. Interface agents will monitor and track the information needs of the user.

3.2 Semantic Inference

The explicit description of innovation related information makes their formal analysis (e.g., goals or goal-action matching among different part of an organization, detecting partially defined and possibly inconsistent goals, actions, or the relationships between them) feasible, which will reduce the human intervention in distributed innovation management. Ontology inference or reasoning, which can improve the efficiency of query and processing of innovation related instance data, will play the role in realizing the alignment analysis among different goals, actions, and results existing in different organizations, projects, and even individuals. Until now, several semantic specification languages, such as RDF [6], DAML+OIL [7], and OWL [8], have been proposed for semantic web. Correspondingly, inference systems for semantic innovation can adopt Description Logic [9], F-Logic

[10], or Horn Logic [11] as its theoretic foundation for the reasoning of innovation related information.

3.3 Semantic Mediation

According to the fact that different organization or project contexts can make statements about the same innovation process using different language, semantic mediation which can make use of concepts and relations defined elsewhere and facilitate the interaction between shared ontologies and different local ontologies is another technical pillar for semantic innovation. Semantic mediation [12] includes ontology consolidation, ontology mapping and alignment. Through ontology consolidation, different parts of a local or several local ontologies are merged as a new shared ontology supporting a specific domain or topic (e.g., innovation management) oriented collaboration. Ontology mapping and alignment allow different terminologies and modelling styles are linked together by creating bridges between separated pieces of knowledge in some domain specific ontologies. These consolidated ontology and the bridges along with domain specific ontologies are then used to perform cross contexts innovation related information search and retrieval (with the ability to query different information sources with different kinds of semantics).

4. SEMANTIC FEATURES

The semantic web is about adding machine-processable semantics to data. The computer can “understand” the information and therefore process it on behalf of the human user. Semantic innovation employs the semantic related technologies to improve the collaboration capability inside or across a networked organization. Every organization or individual use the innovation ontology to describe themselves and published in the trust worthy semantic web. This allows software to process these descriptions, perhaps as part of an automated search engine, to discover information about your and the communities of which you're a member. Semantic innovation has following features:

Goal directed: In the networked organization, different individuals, projects, or department have different goals. To reduce the uncertainty that exists in the networked organization, the innovation ontology gives an explicit and formal specification of the goals in terms of its mission, requirements, strategies and performance indicators. Then the process of formulating goals that reflect a holistic perspective of the organization and horizontal goals matching are done through the supporting of intelligent agent. It allows the organization overall direction to be communicated to not only the human but also the application or machine and empowers them to participate in the innovation process, which assure the high level alignment among these different but overlapping goals.

Action based: Innovation process can be looked on as

the temporal sequence of activities or events that occur in developing and implementing new ideas. To support the correct flow of actions through the various stage-gates of the innovation process, innovation ontologies based semantic mediation and inference are employed in the semantic innovation to ensure a low level consistency within the decision making (about plans and their execution) over time. Ontology-based context aware application is implemented to support the decision-making process concerning selection of actions or plans for projects execution.

Team centered: On the one hand, employee of the networked organization are provided functionality by ontology enabled intelligent agent to capture minimum critical information relating to innovation. On the other hand, the innovation process requires much creativity, problem solving and teamwork, which require a diverse skill base hence any management practice. Each action of the innovation process must have an efficient number of personnel (with corresponding knowledge and skills) assigned to them with a leader being ultimately responsible for action achievement. Inheriting from the class "Agent" of FOAF ontology and extending the attributes of its "Person" class with relevant information of innovation management, the individual member in the networked organization is described in terms of role (academic, engineer or manager), task description (which person is responsible for which goal or action), core competency (knowledge & skills), and relevant area of the task in the innovation process. Through the ontology inference and mediation, semantic matching is realized among goal, capability requirement of action execution, as well as role and the expertise.

Result oriented: The reporting of outcome for all goals and actions are specified extensively and concisely in the innovation ontology, which facilitate the agent which represents corresponding responsible manager to identify all the actions and goals that require immediate attention due to their red status.

5. CONCLUSION

Business competitiveness and sustainability depends on the effective management of innovation. Effective innovation needs to take place within every area of an organisation and by association within key suppliers and strategic partners. Since Semantic Web promises a vision to make information and knowledge machine accessible in the Web rather than displaying them for interpretation by human, a vision of semantic innovation, which exploits the potential of Semantic Web technologies, is presented in this paper to improve effectiveness and efficiency of innovation management in a large networked organization. Several pillar technologies enabling semantic innovation is discussed.

Of course, to actually implement semantic innovation, many other Semantic Web related technologies are also needed to be considered, such as ontology editing, ontology evolving management, semantic annotation, semantic web services, and so on. Until now some of these related technologies are still in their infant stages, more work needs to be done before this semantic innovation vision comes true.

ACKNOWLEDGEMENT

The research work in this paper is supported by the Science Foundation Ireland.

REFERENCES

- [1] Zairi, M., *Best Practice: Process Innovation Management*. Butterworth-Heinemann, Oxford
- [2] Berners-Lee, T., Hendler, J., and Laaaila, O., The semantic web. *Scientific American*, May 2001
- [3] S. Staab, R. Studer, *Handbook on Ontologies*, Springer, 2004
- [4] Dooley L. and O'Sullivan D., "Developing a software infrastructure to support systemic innovation through effective management", *Technovation*, vol. 23, pp. 689-704
- [5] James Hendler, "Agents and the Semantic Web", Available: <http://www.cs.umd.edu/users/hendler/AgentWeb.html>
- [6] Frank Manola and Eric Miller, Editors, "RDF Primer", *W3C Recommendation*, 10 February 2004, <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>.
- [7] Dan Connolly, Frank van Harmelen, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn Andrea Stein. DAML+OIL Reference Description W3C Note 18 December 2001.
- [8] M. Dean, G. Schreiber, S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein, "OWL Web Ontology Language Reference", *W3C Recommendation*, 10 February 2004. Available at <http://www.w3.org/TR/2004/REC-owl-ref-20040210/>.
- [9] F. Baader, D. Calvanese, and D. McGuinness, *The Description Logic Handbook*, Cambridge University Press, 2003.
- [10] M. Kifer, G. Lausen, and J. Wu, "Logical foundations of object oriented and frame-based languages". *Journal of the ACM*, 42(4):741-843, 1995
- [11] Michael Sintek, Stefan Decker. "TRIPLE - A Query, Inference, and Transformation Language for the Semantic Web". *Proceeding of International Semantic Web Conference (ISWC)*, Sardinia, June 2002.
- [12] Ding, Y., & Foo, S. Ontology research and development, Part 2 - A review of ontology mapping and evolving. *Journal of Information Science*, 28(5), 375-388. 2002.

The Analysis of the Difficult Points on Developing E-Commerce of the Western Region in China

Luming Yang¹, Hongli Hu², Cuiping Zhu², Zhuqing1Yang²

¹ Business & Tourism Management School, Yunnan University, Kunming 650091, China

² School of Economics and Finance, Xi'an Jiaotong University, Xi'an 710061, China

Lmyang@ynu.edu.cn, hhl@xjtu.edu.cn

ABSTRACT

From 1999, China started to the project of "Development of Western Region of China" and many preferential policies were issued by the central government. However, after almost 5 years, compared with eastern region, the development of infrastructure is still relatively lower. As to the development of E-commerce, the most typical phenomenon is unbalance which means that the eastern region is much faster than the western because of territorial and economic factors. So it is necessary to get a whole picture and get a clear understanding of problems of current situation of E-commerce in west part of China in order to accelerate it. In this article, the difficult points of E-commerce development in west region are discussed, such as the law issue, infrastructure, information service providers and talents people and some strategies will be given finally based on the current situation of E-commerce in west part of China.

Keywords: Western Development, Electronic Commerce, Internet, Informationization

1. INTRODUCTION

Because of advanced economics environment, E-Commerce of the eastern region has been growing quickly in China. After 1999, China started to the project of "Development of Western Region of China" and at the same time many preferential policies were issued by the central government. From then, the E-commerce in western region has been developing dramatically. According to the reports of CNNIC, from 1999 to 2004, the proportion of the Internet users in the Eastern developed region (Including Peking, Guangdong, Shanghai, Jiangsu, Shandong, Zhejiang) and the western non-developed region (Including Chongqing, Sichuan, Yunnan, Guizhou, Xizang, Shanxi, Gansu, Ninxia, Qinghai, Xinjiang, and Neimeng, Guangxi) has keeping changed gradually. The proportion of the Internet users in the Eastern region occupied in total amount all over the country has been cutting down slowly. But the proportion of the Western region has been growing obviously and the unbalance between Eastern and Western region has improved. Now the infrastructure of E-commerce in the Western region has been improving very quickly and the Internet environment has also been ameliorated gradually. Now almost all kinds of E-Commerce business model and network companies based on business website can be found in west part of China. However, Compared with the Eastern developed region, the level of the infrastructure is still relatively lower. At present, the networks available still can not satisfied with the needs of development of E-Commerce there. If the Western region wants to walk the road of developing its E-Commerce, it is necessary to find the whole picture of existing difficult points of the western E-Commerce. Then, based on the current situation, the proper developing strategies of E-commerce for the Western region will be given.

2. THE EXISTING PROBLEMS ON DEVELOPING E-COMMERCE IN THE WESTERN REGION

Developing E-commerce of western region can promote the economic development greatly and improve its benefit as well. The Western region should grasp the good developing opportunity to make the government and enterprises advance side by side. In the meanwhile, the propagation and study can promote each other to speed up and make the Western region E-Commerce develop healthily through building a good business environment. In order to realize the unity between developing Western region E-Commerce and coordinating the whole western economy, it is necessary to analyze and study the existing problems of E-Commerce in the western region.

2.1 The improvement of market environment of developing E-Commerce in western China

2.1.1 Law issues

E-Commerce is the main content of Network economy, which is a manifest form of market-oriented economy. To some extent, the market-oriented economy means legal system economy. To develop E-Commerce, apart from solving technical problem, it is important to know how to rule the law problems related to the business activities. How to ensure the secure performance of trade on the Internet and other E-Commerce activities is not only a problem of western region, but also is a difficult of problem in all China. It is necessary to restrain some problems related to E-commerce such as proof problem, request of the written form, attestation to trader, consumer, the bank of settling accounts, attestation way of contract effectiveness, working out the effective

electronic contract, the time and place of contract establishment, the rule of exchange behavior between trader and medium, exchange risk and duty problem etc.

E-commerce is a kind of new models of doing business, so there is not a standard legal document in China, even in the world. Although China published its Electronic Signature Regulation of Republic of China (draft) in July, 2002 and made public the Measurement of Management for the Internet Publish and the Temporary Supervision Rules of E-commerce on August 1 2002. But it still can't adapt to the development of E-commerce far and far because part of laws and regulations of electronic trade need to be emended, and some laws and rules need to be added such as the law of Electronic Signature, the law of Electronic Contract etc. So in order to speed up the E-commerce, a fair legal environment is important to protect the consumer rights and secure the trades between the sellers or buyers.

As to the western region, one of vital tasks is familiar with the related laws and rules existing now and to set up additional regulations to improve the law environment of E-commerce there. However, the most import thing is learn to solve the contradictious dispute and issues to safeguard the consumer's legal rights and benefits.

2.1.2 The issues of the infrastructure

In the western region, the start of infrastructure for the networks is later than the eastern China, which resulted in the weak of network applying service. So the networks existing now can not satisfy the need of the development of E-commerce. According to the latest report of CNNIC, most of the Internet users there still use the way of dialing the phone to access to the Internet. The speed of internet is slow and the connecting time is longer too. Although most of new buildings there lay the internet access line into doors while they build them, most of them can not provide any practical content and services except the basic facilities for the community. Especially, the breadth grows slowly because many real estate developers just take the breadth network as a selling spot. Most Internet users can't understand the technical index of breadth network, speed and network service in community as well as some developers. To some extent, those facts restrain the development of E-commerce in the western region.

In western region, the universal rate and the applying level of the computer are low. For example, many people there don't know how to use computer and the Internet. At present, the applying level of the computer in many enterprises is behind, meanwhile, many of them lack the information system to support their work. Even for the enterprises who introduced computer and the Internet into their companies, the main uses of computer are word process and calculation. Most of business activities in companies, such as product, supply, market, management of human resources are lack the support of

information technology. Such undeveloped management only enable the computer to simulate the operating process of original handicraft simply, and make it difficult to realize E-commerce system and increase the invest cost as well as cut down the rate of E-commerce returns.

As the third-part authority provider and fair attestation department, CA is an important part which ensures the safe activity of E-commerce. However, an effective supervisory of CA and the safety system of information are lack or weak. Because the rights and duties of CA can offer a person's true ID and use higher encrypt technology, it also related to the function of government department and the safe problem of the network. Up to the present, many organizations which take CA as their main business started because there is no integral measurement for CA from stating to operating in China. However, many CA owners may lack the capability and reputation, and technology. In case that the CA owners lost their business, it will be a big problem to handle the user's data.

Electronic payment system of the bank hasn't been established completely in the western region. The payment system of E-commerce has been set up by China Telcom, China Agriculture Bank, Chinese Bank, Merchant bank and other domestic bank etc. But because the system is not perfect, only half of the users who shopping via the Internet pay online. Another half still adapt the cash payment means on delivery.

As for the logistics, the modern logistics and its socialization are still staying in the stage of conception. In the logistics enterprises, it is obvious shortage on the information application, overall planning and the fine turns organization, as well as the ability of the management. On the other hand, the level of the specialization and technique in logistics system of the western region are all lower that they can't adapt to the requests at all on developing information flow and funds flow of E-commerce.

In summary, the government of the western region should pay attention to the infrastructure of E-commerce, and increases the investment into the construction of hardware facilities in order to solve the existing problems.

2.2 The development of information industry in the western region can't be satisfied with the need of the E-commerce.

2.2.1. The current situation of information system in western region

The developed countries such as Europe, American and Japan, the integration of the three greatest systems including SCM (supply chain management), ERP (business enterprises resources), and the CRM (customer relation management) have been ripe

basically. From MRP, MRPII to ERP, SCM, CRM, its development has gradually experienced for more than ten years. But the western region sprang over the developing process gradually, they introduce advance management thoughts and technique directly, most adapt the difference E-commerce business mode in many large enterprises, only a few make a success. This states that the western region business still can't break away from the traditional business mode in the course of carrying out E-commerce at the present stage. The information turns of the western region business is still in the stage of the beginning and exploring. They should set up electronic commerce terrace according to the present condition, all kinds of technique and developing conditions in the present business flows.

2.2.2 The information service industry in the western region

The software and hardware products mainly depend on the products of the foreign companies because the manufacture of the Chinese information equipments and

the software industry can't satisfy with the demands of soft-hardware products in the course of the development of the information and E-commerce in the western region .In addition, the actual strength of economy in the western region is poor.

2.3 Lacking strategic planning on developing E-commerce because of lower information level in western region business

On account of the proportion of the websites amount is one of the index to measure the level of the information, according to the reports of CNNIC, he proportion of the registered websites under CN in South west and northwest has increased a bit in recent years, but at the same time the proportion of registered websites in the North, the South, the East and Northeast has been occupying more than 85%, this reflects the gap between western and eastern region in the development of information infrastructure.

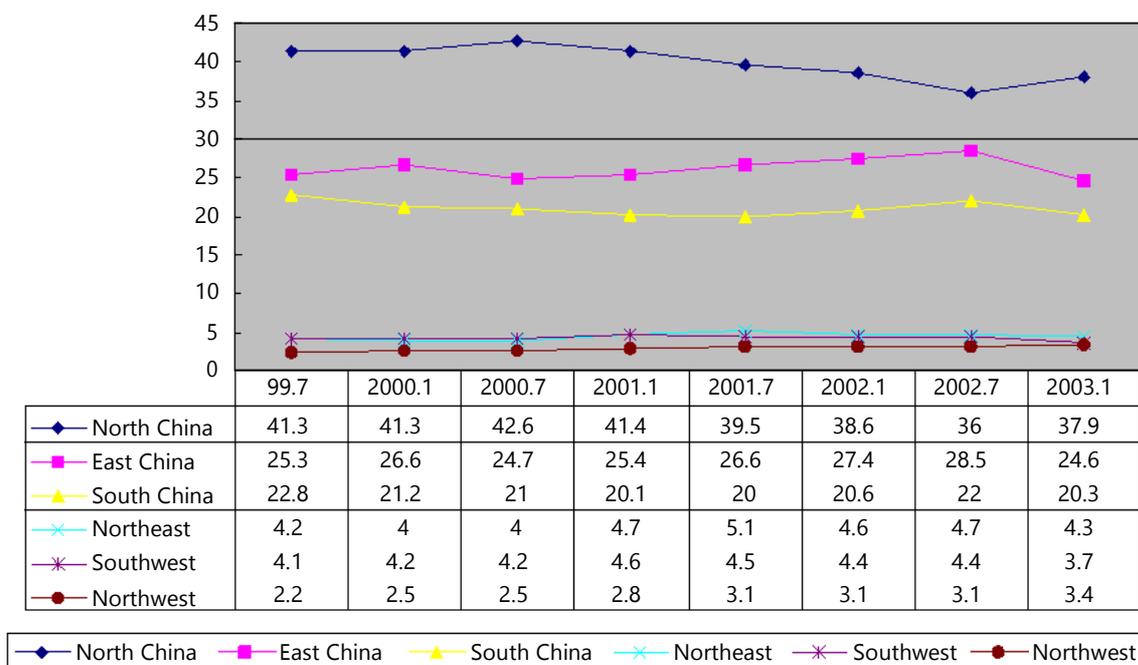


Figure : The registered area name and area distribution under CN

There are 2000,000 enterprises in the western region in China, and 10,000,000 families in the town, but the proportion of Internet access is relatively smaller. There is a very great margin between western and eastern region. According to the study of national statistic center, the ability of the information in the western region not only fall behind the developed regions such as Peking, Guangdong, Shanghai, but also fall behind other region such as Hubei, Henan and Liaoning as well. The development in domestic is very imbalance. There is a big margin in Internet users and computer owners between the western and eastern regions. Because the computer application in the western region starts lately,

at present, the construction of the business information in the western region is low, especially in the aspect of Network. The most only registered local area name and announced the basic business information on Internet. About Internet applying, many enterprises were still in the stage of exploring. Most business haven't realize the information turns, the means of information process and handle lags behind, so does the ability of information processing. So the demand of business for E-commerce is still at the primitive stage. Though the Internet users in the western region have increased doubly, the number of the Internet users only 1600,000.

Lacking the long-term planning of the E-commerce for many enterprises in the western region, even though many business begin to give energetic impetus to E-commerce, the invest returns to build their website is less, and prestige is also lower. The business website constructed themselves can get the purpose of publicity, but it is unknown that how many people can buy commodity on Internet because of influencing under the scales of business and the position of the website in the same industry. The repeated construction of the website will become the disadvantage for sharing Internet resources with the same industry. Therefore, only by working out the developing E-commerce plan consequently, the business in the western region will have a long-term development.

2.4 The issue of lack talent people

The talented people that the E-commerce needed can be divided into three levels. The first is the support layer of technique service. The staff in the first layer has mainly pay their attention to the technical aspect of E-commerce such as Network building, system management, main page creation, developing program etc. In China, the amounts of programmers is enough, the means of education can't basically meet the need of development. The first grade talented persons on system management are still scarce, especially in the western region. Lacking of web manager, Unix system manager, and the payment of training is still expensive. The level of many system managers is low, which only can be used to do some daily upholding work. They can do little job on higher management like safety and performance adjusting, for this reason, many systems in the western region exist hidden trouble. The second is the manage layer (middle-level manager, financial staff, middle and lower government staff etc). The secondary staff is the largest and they are the personnel of the business or section and the direct beneficiary of E-commerce. They don't care about the technical details about E-commerce. It's enough for them to do their work under the electronic way. But to a large extent, the E-commerce education neglected the needs to this layer. The third is the higher level of the manager. To the third level of structure, that means higher manager, it is not considered on the whole by E-commerce education. It is a very key link to push the popularization and training of E-commerce quickly if the E-commerce can grow smoothly in the western region in China. And go a further step, paying attention to attract the talented person, train and educate them are the key factors too.

2.5 The issue of the role of the local government

Compared to the developed region, the information technique in the western region fall behind, it will contribute to introduce technique and promote the development of E-commerce if the government can participate in it. The macroscopic planning and guide of the developing E-commerce will contribute to coordinate

the developing balance and the benefits among various areas. But the main role of the E-commerce development still belongs to enterprises. The government should mainly have the eyes on the macroscopic planning and guiding, not interfere the business, on the contrary, they should put the power to it and obey the international standard of E-commerce. At present, the western region is faced with the following environment such as a part of enterprises can't be separated from the government, the credit system is incomplete and the local protectionism is very severe too. Facing this kind of situation, how can the government department work well and adjust and guide positively will play an important role to the development of the E-commerce in our western region.

2.6 The issues of conception and consciousness

Because the development of E-commerce is still on its early stage, the consumers still remain the traditional consume habits of "see, touch, hear" and this shopping habits is deep-rooted, the majority still remain the shopping principle of "delivery when pay". Most consumers suspect the security of pay, quality of goods and hesitate to shop online because they can not see and touch the good via the Internet. At the same time, the density of the physical market is very large in the city structure in western region, so the consumers can buy anything what they like as soon as they go out. Compared with the E-commerce that has unperfected in the current logistics system, on the contrary, the consumers feel that the traditional shopping ways is more convenient. This makes the advantage of the "convenience", "fast", and "a good function of alternativeness" that the E-commerce depends on can't be reflected, to some extent, it will restrain the development of E-commerce in the western region.

On the other hand, there is a little information that Internet can directly bring economic benefits for business, but the general information is more. They often can't find profit about what they imagine on Internet and this will discourage the initiative of business.

3. The strategies on developing E-commerce in the western region

Developing E-commerce is a social system project which relate to the government, business, financial industry and individual as well, and it needs the coordinating of the social environment, overall planning gradually and cultivating actively in the course of construction of the basic facilities. Therefore, the western region should grow according to the whole designs and goals on developing E-commerce.

The first, having the positive attitude, the government will play an important role to push the development of E-commerce. Particularly, on establishing Internet basic facilities, building common terrace, forming electronic attestation and the standard of electronic payment, at the same time, they should create a good support

environment of technique, finance, legal system etc. At present, the government is issuing the law related to the E-commerce, regulations and management methods to ensure it to be developed healthily. With the development of the western region, the local government should be based itself on the actual states and worked out the developing plan of local E-commerce. Besides this, they should coordinate the basic facility construction of the Network and communication, they also should try their best to save funds and reduce the repeated funds to minimum. In the meanwhile, giving the preferential policy of the tourism, industry and finance, tax revenue, communication, financial examine, funds and science study to the mature region and unit to encourage experimental unit and other invest businessman to invest strongly, At the same time, They should develop and study the E-commerce so that they can further explore the road to develop E-commerce.

The second, introducing the medium to enlarge the publicity dint degree of the E-commerce and encouraging scholar and business managers to study and publicize the E-commerce in order to get a good ready for developing E-commerce in the western region. It is a good way to enhance the correct knowledge of E-commerce to the business managers in a discussing way at a proper time. It is necessary to organize the business managers to study theoretical knowledge systematically. It also can organize them to put into practice where the E-commerce can be carried out well. It needs to make a long effort in the whole society to change the shopping habit of the public and establish the reliable system of consume as well as the operating ways of market.

The third, we should attract and educate the talented person to adapt the development of E-commerce. This is a basic and difficult point. It is difficult to attract the higher talented person because the western region is located in remote place, the level of economy and productive force falls behind, and the living standard is poor, too. So we should try our best to take the feasible measure to attract those who are willing to contribute themselves to the development of the western region. At the same time, it is important to provide a good working environment for them to enable them elaborate their talent. At the same time, it is important to pay attention to training the existing talented persons.

The fourth, with the market and performance as a target to enhance the of business management when the construction of the information is proceeding. Some famous enterprises in the world can keep the competitive advantage because they think highly of coordinating all kinds of productive forces and enable the links of decision, production and market to operate efficiently through gathering every kind of technique together. So the construction of business information turns can't only consider the business itself, they should still consider to connect with other unit through Internet including some

issues about the sign, confirm, pay, goods arrival of national contract etc. That is to say, the enterprises should improve their modern management level and establish internal information management system of the computer to enable itself to adapt the development of E-commerce in the future. The local government can specify technique organization or scholarship groups to offer consult and technical support for internal information management system of the computer and the development to ensure the business to get a good development in the future.

The fifth, it is important to draw lessons from the domestic and international advanced experience and take the current situation as a development base. Up to now, about the development of the E-commerce, there are many experiences in the advanced country and domestic examples of success for the business in the western region to learn from. For instance, the three greatest Bottle Neck problems on Internet attestation, ration system, electronic payment that had been restrained the development of E-commerce in China, now have already solved some to some extent. Some Internet businesses are also exploring a little successful experience in their actual operation. Combined the condition by themselves, the western region enterprises should use the successful experience and study carefully about which business and areas can fit to develop E-commerce in order to explore a new road for the western region to develop E-commerce and ensure it to have a good beginning in the western region.

The development of E-commerce is a systematical project which concerns with the corporate, consumers and government as well. The development of E-commerce in the western region is in its early stage and many difficult points are still there, so it will be a long way to go to catch up with the step of E-commerce in eastern China. It need the whole society makes a joint effort to coordinate and cooperate to promote E-commerce of the western region.

REFERENCES

- [1] Yearbook of Chinese electronic commerce, 2002
- [2] Yangluming, The measure of E-commerce development of China joins the WTO, The Ideological Front, 2001
- [3] Ying Liannian, Exploring the new management mode of business E-commerce organization, Commerce research, pp 265-269, 2003
- [4] Yuquan, Yinxinyu. 2002 E-commerce ten key words, Chinese electronic commerce, 2003
- [5] Dongbigeng, Rebuilding the commerce flow in all society, Chinese electronic commerce, vol. 2, 2001
- [6] Caiyanbing, Jiasutong, E-commerce--walks into the digital figure turns in business ages, Intelligent property right publishing house, 2000
- [7] Liguie, Thinking highly of the development in the development of the western region, vol.6, 2000

The Business Process Investigation in the Perspective of Customer Value

Ray Tsaih^{1*}, Wanying Lin², Miaoju Hung³, Yaling Cheng⁴

¹Department of Management Information System, National Chengchi University, Taipei, China

²Department of Accounting, National Chengchi University, Taipei, China

³Department of Management Information System, National Chengchi University, Taipei, China

⁴Department of Management Information System, National Chengchi University, Taipei, China
{tsaih, m92027, m92015}@mis.nccu.edu.tw, wanying@nccu.edu.tw

ABSTRACT

To identify the customer value provided from the business process and to seek additional opportunities to fortify a firm's competitive advantages, we explore the value chain mentioned in (Porter, 1985) at a greater extent. The study results in a framework where an information-processing infrastructure is added to the value chain concept to provide a basis for the business process investigation. The essential concept of this framework is the **process-wide information organism (PWIO)** approach for describing and examining mechanisms of managing the information flow that contributes to customer value creation.

Keywords: process investigation, customer value, value chain, linkage.

1. THE PROCESS MANAGEMENT AND THE VALUE CHAIN CONCEPT

A business process consists of a series of related business tasks to be performed. An ordering process and a transaction-based revenue cycle are commonly seen business processes. Business processes create and deliver a value proposition for customers through the (value-added) product or service¹. Thus, the performance of business processes is a leading indicator of subsequent improvements in (customer) value and financial outcomes. To well manage business processes and to seek (additional) opportunities to fortify a firm's competitive advantages are important issues nowadays.

In the literature, there are many studies of (business) process analyses. In terms of describing and analyzing processes, these studies can be grouped based on two distinct viewpoints for different targeted readers: one focuses on the technique issues and the targeted readers are engineers (for example, the IDEF0), and the other talks about a firm's strategy and the targeted readers are executives (for example, the value chain concept). Each viewpoint has its predicaments.

More specifically, the Process Management (PM) method suggested in (Elzinga, *et al.*, 1995; Stefano and Giovanni, 2003) is set out in the following six basic stages:

- (1) *Preparation*. Defining the crucial factors for business success and a firm's basic strategic targets; designing the organization structure (for example, steering committee, PM teams, etc.) for PM.
- (2) *Process selection*. Identifying the process to be analyzed and improved.
- (3) *Description*. Agreeing on a common definition of the process by documenting workflows.

(4) *Quantification*. Defining the process in terms of cost, quality and time performance.

(5) *Selecting improvement opportunities* based on acquired knowledge of the process, information about the crucial factors for business success, and of company targets.

(6) *Implementing improvements* that have been identified.

The cycle can be repeated downstream of the improved process, in order both to select new processes and to seek further improvement opportunities within the same process. Integration Definition for Function Modeling (IDEF0) stated in (Sandie, 1997) can serve stages of *Description*, *Quantification* and *Selecting improvement opportunities* of the PM method. IDEF0 is a model used to present structured information about the functions performed by an organization and has been proposed to analyze and identify process improvement opportunities. In addition, simulation modeling proposed in (Hlupic and Robinson, 1998) could offer great potential in analyzing business processes through dynamically modeling different samples of parameter values such as arrival rates or service intervals which can help in discovering process bottlenecks and investigating suitable alternatives. Some drawbacks in IDEF0 and simulation modeling are the excessive usage of technique and the inability to simplify the analysis.

On the other hand, some studies claimed that business process analyses should be based upon the customer value and a firm's strategic viewpoint. For instance, Porter (1985) stated that the issue should be the value, the amount customers are willing to pay for what a firm provides them. To discover a firm's competitive advantages, Porter proposed the value chain concept to analyze buyers, suppliers and a firm via disaggregating them into discrete but interrelated activities from which (customer) value stems. Furthermore, Kaplan and Norton

¹ Such service includes the ones for internal and external customers.

(2001) observed that organizations focusing on strategy and customer value tend to be more successful. With a clear picture of business direction and customer value propositions, organizations can deploy (or redirect) its resources at implementing strategy initiatives. In addition, the availability of knowledge of information system infrastructure and applications that align information capital to the strategic business processes is nowadays required to support the implementing strategy initiatives. As such, the expected outcomes are more attainable. Unfortunately, these studies did not release any solid methodology for examining the customer value delivered from the business process.

For instance, each value chain in Porter (1985) was designed to fulfill a set of value-adding activities. Across activities, there are linkages where information flows allow *optimization* or *coordination* to take place. Linkages are “relationships between the way one value activity is performed and the cost or performance of another.” (Page 48) However, Porter stated that it is much more difficult to recognize and manage linkages than managing value activities themselves, and the ability to recognize and manage linkages “yields a sustainable source of competitive advantage.” (Page 50) Without a solid knowledge for recognizing and managing linkages, it is difficult to effectively examine the customer value provided from the business process that covers a subset of activities.

These unsolved predicaments motivate this study. This study tries to explore the customer value delivered through the business process. For such an exploration, this research proposes a framework where an information-processing infrastructure is added to the value chain concept to provide a basis for the business process investigation. The essential concept of this framework is the process-wide information organism (PWIO) approach for describing and examining mechanisms of managing the information flow that contributes to customer value creation.

In the following sections, we will first introduce the PWIO approach, followed by the presentation of the business process analytical methodology. Promising managerial applications and future work are given in the last section.

2. THE PROCESS-WIDE INFORMATION ORGANISM (PWIO) APPROACH

Every firm maintains a variety of groups of functionally interrelated contact points (CPs) which work together for performing activities of the value chain mentioned in (Porter, 1985). The value chain concept is proposed to analyze activities a firm performs, and how these activities interact to procure competitive advantage. After receiving request for service from other (intra- or inter-organizational) activities, each activity plays the dual role as a server of value-adding provider (executing

business tasks for the client) and, if necessary, as a client (releasing information for asking for support from other activities). In other words, there are two kinds of information-processing infrastructures practiced in performing a business task: (1) CPs to process information and conduct the corresponding business task, and (2) information channels (ICs) to carry the information flow.

The practice of each CP processes information and undertakes business tasks for value-adding. In the CP, there are mechanisms capable of (1) receiving the information delivered from other CPs, (2) executing business tasks based upon the received information, (3) feeding back the corresponding information, and/or (4) releasing information to the supporting CPs. Between two different CPs, a client and a server, the (intra- or inter-activity's) information flow should pass forward or backward through the IC seamlessly to request or provide a service. Thus, each CP accompanies with the inbound and outbound ICs for carrying the in-coming and out-going information, respectively.

Hereby, a business process consists of a series of related business tasks that the practice of a group of functionally well-connected CP/IC is designated to perform.

In practice, a staff, an ad hoc team, or an information system could act as the practice of a CP. And a memo, a post mail, a phone call, a face-to-face meeting, a teleconference meeting, an intra net, an email, a (beeper or wireless) message, or a web information system could act as the practice of an IC. Furthermore, there may be multiple practices of a CP/IC. In brief, the practice of the CP/IC is in one of the following forms: manual, automatic, or parallel (have both manual and automatic practices simultaneously).

The practice of each CP/IC should create value (in the form of lower costs or improved performance) for the client CP. Difference in practice of a CP delivers various level of quality of coping with the corresponding business task, which differentiates the outcomes of its value-adding. For instance, an experienced manager can deal with unexpected business scenarios better than a decision support system and a junior manager. Difference in practice of an IC also differentiates the outcomes of the associated CP's value-adding. For instance, a phone line at a time epoch can serve only one customer, but the audience served by an Enterprise Information Portal at any time epoch may be more than one. The costs associated with different practices of each CP/IC may vary as well.

The discussion of the ICs explores the identification and management of linkages among value activities because, as mentioned in Porter (1985), exploiting linkages requires information or information flow which allows optimization or coordination to take place. Each IC identifies the existence of a linkage between two CPs,

and to identify the degree of integration of CP/IC is to explore the management of linkages. A better integration of CP and IC can render a better optimization or coordination happened in the information flow to leverage linkages between (value) activities. For instance, a firm provides an email practice in the (outbound) IC to customers, but some customers are incapable of receiving emails. In other words, practices of this very (inbound/outbound) IC between two CPs are asymmetric. The asymmetry existing in practices of the (inbound/outbound) IC between two CPs would impede the effectiveness of communication, the performance of the task, and thus the corresponding customer value.

The collective quality of information and the smoothness of the information flow associated with a business process are determined by the practices of all involved CP/ICs. The higher the degree of the collective quality and smoothness of information flow is underlying a business process, the larger the value a company can derive from these connected CP/ICs. However, a poor practice of an individual CP deteriorates the collective quality of information and the smoothness of the information flow in a business process.

The PWIO of a firm embodies all practices of all CP/ICs of a business process. Business processes that are critical to business success are expected to proactively responsive to the dynamic business environment. As such, each group of functionally well-connected CP/IC is adaptive in term of its functionality and practice. This adaptation leads to an analog between the PWIO and an organ in the living body.

Collectively, a firm maintains a PWIO for a business process. Via the practices on all CP/ICs in the PWIO, the firm handles not only the corresponding (transaction and non-transaction) information flow, but also performs all the involved business tasks. Furthermore, a firm uses its PWIO to do all the involved business tasks in a whole, not separately, for doing business. Whether a firm has a good performance in the (decisive) business process in value-adding for its customers depends on how well it manages the corresponding practices of CP/ICs of the associated PWIO.

In closing, the PWIO approach stems from the value chain concept and provides a solid description of the business process without too-much technique.

3. THE BUSINESS PROCESS ANALYTICAL METHODOLOGY

The proposed business process analytical methodology basically follows the first four stages of the PM method. Executives practicing the art of strategy must identify the decisive business processes that are the most essential for creating and delivering the differentiating customer value propositions. And the identification should be aligned with the vision of the firm. The revenue cycle, for

instance, is a decisive business process.

For each decisive business process, the following steps are proposed for accomplishing stages of *Description* and *Quantification* of the PM method: (1) to develop the corresponding business process table, (2) to develop the associated practice tables, (3) to complete the performance measures of CP/IC shown in practice tables, and (4) to integrate the summarized performance measures derived at the step (3) into the decisive business process table. The detail of each step is presented below.

(1) Developing the business process table: There are four columns in a business process table: responsibilities, functions, objectives, and performance measures. At the current step, only the responsibilities and functions columns are determined and filled in.

(2) Developing the associated practice tables: The practice table is to explore the practices on all associated CP/ICs for each function listed in the business process table. There are seven columns in the practice table: the practice of the associated CP, the job requirement of the CP, the practice of the associated inbound IC, the job requirement of the inbound IC, the practice of the associated outbound IC, the job requirement of the outbound IC, and the associated objectives and performance measures. At the current step, except the column of the associated objectives and performance measures, all the other six columns are determined and filled in. The corresponding practice of the other connected CP is shown in the parenthesis. The redundant CP/ICs and less important CP/ICs should be omitted.

(3) Completing the performance measures of CP/IC shown in all practice tables: Management cannot take place if the object needed to be managing cannot be measured. The measurement however would be impossible without a good description of the object. Therefore, for each practice table, we first determine and fill in the associated objectives of CP/IC, and then assign the performance measures correspondingly.

In each practice table, only the summarized performance measures are listed, and there are no detailed performance measures, the ones corresponding to the connected CP. For instance, when the CP in the sales department asks the corresponding CP in the production division for the backing up order information, the summarized performance measure is the process time of asking backing up order, and the detailed performance measures are process times of checking production schedule (performed by the CP responsible for production scheduling) and raw materials (performed by the CP responsible for raw materials warehousing).

(4) Integrating objectives and summarized performance measures associated with all practice tables into the columns of objectives and performance measures

appeared in the business process table, respectively.

In brief, the former two steps correspond to the stage of *Description* via following the PWIO approach presented in Section 2. The latter two steps correspond to the stage of *Quantification* of the PM method. In the literature, there are bodies of performance measures that we can borrow from for accomplishing the step 3. For instance, we can reference the measures of IT/IS's profitability, productivity and customer value (cf. Hitt and Brynjolfsson, 1996; Laudon and Laudon, 2000, pp. 353-360; Devaraj and Kohli, 2002, pp. 6-8; DeLone and McLean, 2003) to get the candidate performance measures of the practice of each automatic CP/IC. The candidate performance measures include, but not restricted to, an amount of errors, severity of errors, response time (process time), rate of tasks performed, the reduction in process/search costs, sales growth, market share, customer satisfaction/loyalty (repeat purchases), old customer retention, new customer acquisition (growth in new sales), and customer profitability contribution. There are also references in the human resources (cf. Bohlander, Snell, and Sherman, 2001), for the measure of the practice of manual CP/IC. But to measure the practice of manual CP/IC is much more difficult than to measure the practice of automatic CP/IC since the former involves not only the productivity and the profitability, but also the personality, including intelligence, experience, creativity, and (sometimes) ambition. A simplest way is to adopt the same candidate performance measures to both automatic and manual CP/IC.

Table 1 gives the illustration of a business process table corresponding to the revenue cycle of a specific firm² and Figure 1 gives the corresponding depiction. Tables 2 and 3 give the illustration of practice tables for functions of developing new customers and asking for backing up order shown in Table 1, respectively. In Figure 1, the parenthesis shows the corresponding practice of the connected CP.

4. PROMISING MANAGERIAL APPLICATIONS OF THE PWIO AND FUTURE WORK

Business objectives and performance measures are rather common to all companies for planning, controlling, and performance evaluation purposes. However, the

alignment of objectives and performance measures without targeting at business direction could be detrimental to organizational performance (Kaplan and Norton, 1997). Thus the development of objectives and measures should depend on business direction/strategy of the firm. As the proposed business process analytical methodology is based on exploration of information flow underlying value chain activities, the business direction, building up customer value, is clear. With the clear business direction, a strong coherence of process objectives and performance measures buttresses implementations of customer value-oriented business processes. In this perspective, the analysis resulted from the PWIO approach helps the analysis of the implementation of business strategy.

It is worthy of noting that, for carrying the information, the practice of IC alone is always a necessity in the business environment, and nowadays gradually becomes a utility that is rather irrelevant to the competitive advantage. Furthermore, since the effectiveness of the integration CP/IC and thus quality of business processes determine how value get created and sustained, the proposed business process analytical methodology provides a venue to the business process evaluation.

Some knowledge extended from the PWIO approach, for example, the one in Figures 1, also depicts the level of integration of practices of CP/IC in forms of manual, automatic, or parallel. The automatic practices of CP/ICs correspond to the information system infrastructure and applications. In a sense, the information systems evaluation can be done implicitly via examining the extent and the importance of the corresponding automatic practices of CP/ICs in the whole (decisive) business process. Thus, the PWIO approach also provides a means of obtaining the knowledge of information system infrastructure and applications that align information capital to the strategic business processes. A future work is to strengthen such a sense and complete the associated studies.

ACKNOWLEDGEMENT

This study is partly sponsored by Institute for Information Industry, Taiwan, China, Grant No. 92-0524

² The study of this revenue cycle is based upon the case study of a project sponsored by Institute for Information Industry, Taiwan, China. For details of all practice tables please send a request to the corresponding author.

Table 1: The business process table for revenue cycle of some firm

Responsibility	Function	Objectives	Performance Measure
1. Developing and contact customers	1-1. develop new customers 1-2. contact customers	<ul style="list-style-type: none"> • develop new customers and new sales • prompt response • increase customer satisfaction 	<p>growth in new sales</p> <p>response time</p> <p>customer satisfaction</p>
2. Processing sales order	2-1. take order 2-2. approve credit 2-3. check inventory availability 2-4. ask for backing up order 2-5. approve / reject order	<ul style="list-style-type: none"> • prompt in taking order • reduce process time of approving credit • reduce process time of checking material • reduce process time of checking product availability • prompt in checking backing up order • precision in taking order • precision in inventory record • accuracy in raw material record 	<p>process time of taking order</p> <p>process time of approving credit</p> <p>process time of checking material</p> <p>process time of checking inventory availability</p> <p>process time of checking backing up order</p> <p>error in taking order</p> <p>error in inventory record</p> <p>error in raw material record</p>
3. Picking, packing & shipping	3-1. pick& pack products 3-2. ship products	<ul style="list-style-type: none"> • prompt in preparing for shipping • precision in picking & packing order • accuracy in invoicing & shipping 	<p>cycle time for preparing for shipping</p> <p>error in picking& packing order</p> <p>error in invoice & shipping</p>
. Billing & maintaining accounts	4-1. update accounts 4-2. bill	<ul style="list-style-type: none"> • timely update customers' credit record • reduce process time of inquiring credit • screen customers with bad credit • accuracy in customers accounts • precision in billing 	<p>interval of updating credit record</p> <p>process time of inquiring credit</p> <p>amount of bad debts</p> <p>error in customers accounts</p> <p>billing errors</p>
5. Collecting cash	5-1. remit & deposit	<ul style="list-style-type: none"> • timely collection of account receivables 	<p>account receivable turn over</p>

Table 2: The Practice table for developing new customers

Practice of the CP	Job requirement of the CP	Practice of the inbound IC	Job requirements of the inbound IC	Practice of the outbound IC	Job requirement of the outbound IC	Objectives (Measures)
Salespeople in Sales Unit 1 of Sales department	develop new customers in Taiwan	teleconference meeting, mail, phone call (from CEO or secretary of CEO)	bring in information			develop new customers and Increase rate of customers acquisition (rate of customers acquisition)
Salespeople in Sales Unit 2 of Sales department	develop new customers in Mainland China					
Salespeople in Sales Unit 3 of Sales department	develop new customers in other area					

Table 3: The Practice table for Asking for backing up order in Sales

Practice of the CP	Job requirement of the CP	Practice of the inbound IC	Job requirement of the inbound IC	Practice of the outbound IC	Job requirement of the outbound IC	Objectives (Measures)
assistants in sale-supporting Unit of Sales department	ask the Warehousing department and Production department for backing up order. If the due date cannot be made, then acknowledge customer the new due date.	<ul style="list-style-type: none"> • ERP information system (from raw material inventory information system in Warehousing department) • ERP information system (from production management information system in Production department) 	bring in the result of backing up order (from Warehousing department and Production department)	<ul style="list-style-type: none"> • ERP information system (to raw material inventory information system in Warehousing department) • ERP information system (to production management information system in Production department) • phone call, email (to customers) 	<ul style="list-style-type: none"> • deliver the asking for backing up order (to Warehousing department and Production department) • acknowledgement of whether the production is necessary (to customers) 	<ul style="list-style-type: none"> • reduce the process time of asking for backing up order (process time of asking for backing up order) • reduce the process time of checking raw material (process time of checking raw material)

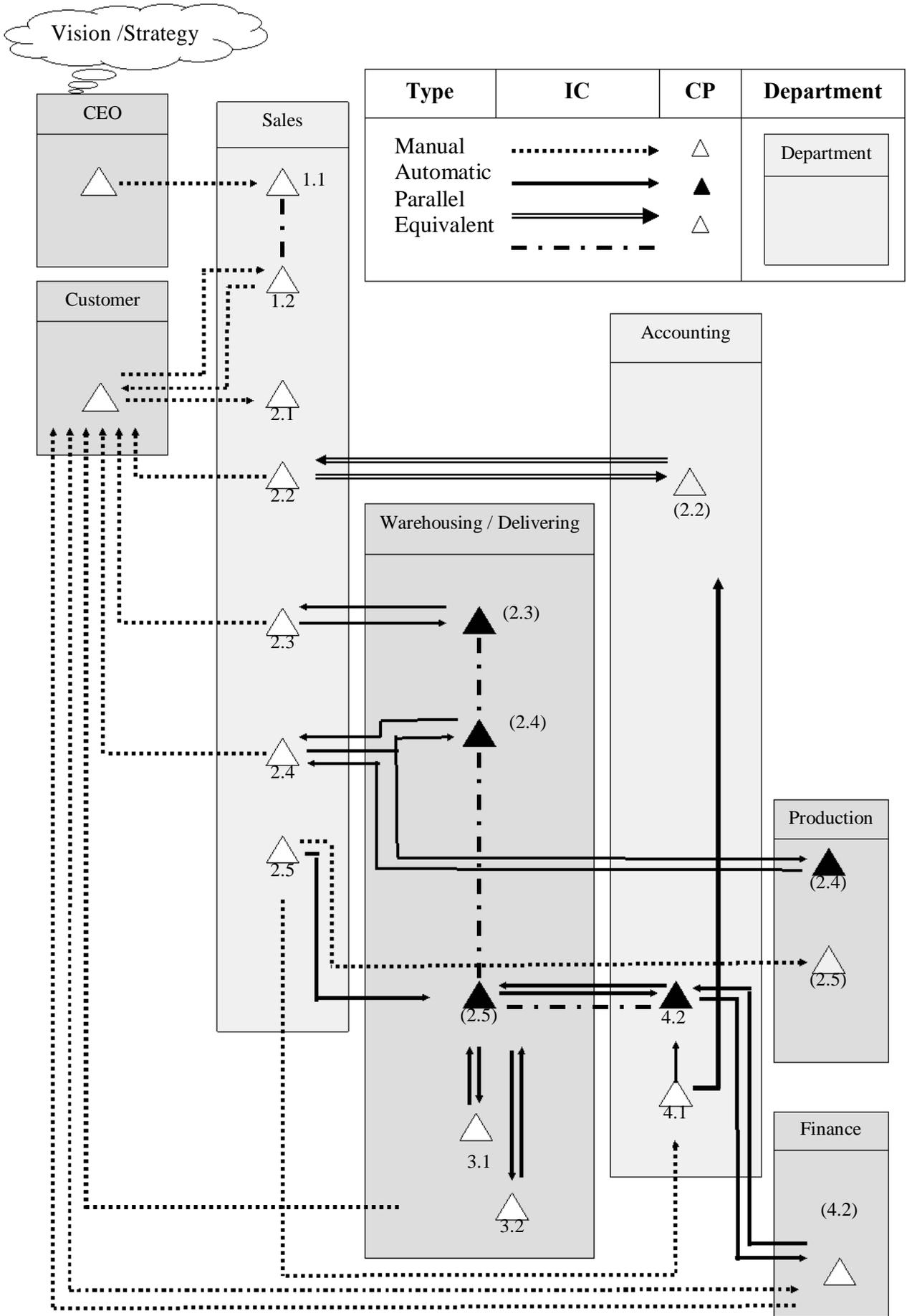


Figure 1: A graphical illustration of the CP/IC deployment in the revenue cycle shown in Table 1 derived from the PWIO approach.

REFERENCE

- [1] Bohlander, G., Snell, S., Sherman, A., *Managing Human Resources*, 12th Ed. South-Western College Publishing, Cincinnati, Ohio., 2001
- [2] DeLone, W. H., McLean, E. R., "The DeLone and McLean model of information systems success: a ten-year update", *Journal of Management Information Systems*, Vol. 19, No. 4, pp. 9-30, 2003
- [3] Devaraj, S., Kohli, R., "The IT payoff: measuring the business value of information technology investments", *Financial Times Prentice Hall*, New Jersey., 2002
- [4] Elzinga, D. J., Horak, T., Lee, C. Y. and Bruner, C., "Business process management: survey and methodology", *IEEE Transactions on Engineering Management*, Vol. 42, No. 2, pp. 119-27, 1995
- [5] Hlupic, V., Robinson, S., "BUSINESS PROCESS MODELLING AND ANALYSIS USING DISCRETE-EVENT SIMULATION", In Medeiros, D., Watson, E., Carson, J. & Manivannan, M. (Eds.), *Proceedings of the 1998 Winter Simulation Conference*, pp. 1363-1368, 1998.
- [6] Hitt, L. M., Brynjolfsson, E., "Productivity, Business Profitability, and Consumer Surplus: Three Different Measures of Information Technology Value", *MIS Quarterly*, Vol. 20, pp. 121-142., 1996
- [7] Kaplan, R. S., Norton, D. P., *The balanced scorecard*. Harvard Business School Press, Boston, Massachusetts, 1997
- [8] Kaplan, R. S., Norton, D. P., *Strategy focused organization*, Harvard Business School Press, Boston, Massachusetts., 2001
- [9] Laudon, K. C., Laudon, J. P., *Management information systems: organization and technology in the networked enterprise*, 6th Ed. Prentice-Hall, Upper Saddle River, New Jersey., 2000
- [10] Porter, M. E., *The value chain and competitive advantage*. In: *Competitive Advantage: Creating and Sustaining Superior Performance*, Free Press, New York., 1985
- [11] Sandie K., "Putting your IDEF0 model to work", *Business Process Management Journal*, Vol. 3, No. 2, pp. 151-161, 1997
- [12] Stefano B. and Giovanni B., "Risks and opportunities of the new ISO 9001 certification", *Business Process Management Journal*, Vol. 9, No. 2, pp.149-169, 200
- [13]

The Internet and the Organizational Adoption of Electronic Business: A Research Agenda

Sanjiv D. Vaidya, Merlin Nandy

MIS Group, Indian Institute of Management Calcutta, Joka, Kolkata 700104, India
{sdvaidya, merlin}@iimcal.ac.in

ABSTRACT

The late nineties witnessed adoption of the Internet by organizations, on a very large scale. However, not every organization did experience an equally deep penetration of the Internet in its business. It was followed by the dot-com bubble burst. Many researchers have studied the phenomenon of organizational adoption of the Internet and the associated success factors. However, the available literature is fragmented in nature and does not take into consideration certain fundamental concepts. Therefore, there is a need for an enhanced understanding of the phenomenon. It will enable executives to plan for and implement Internet adoption in a more effective manner. This paper discusses the literature and brings out the inadequacies in the current findings. Subsequently a framework is developed to indicate directions for future research.

Keywords: Internet, IS adoption, E-Business adoption, E-Business drivers

1. INTRODUCTION

Information technology (IT) has been a part of virtually every business, particularly in the developed world, for the last two decades. The advent of the personal computer (PC) in the early eighties brought in a quantum jump in the scale of adoption of IT by organizations. The late nineties witnessed yet another wave of IT – the widespread adoption of the Internet by organizations, on a very large scale. Many researchers, from the fields of business strategy and information systems alike, have studied the phenomenon of adoption of the Internet by organizations in general [Stephan 2002], and the Internet boom and bust in particular. It has been observed that at one extreme of the spectrum of adoption of the Internet, organizations just put up a website with skeletal information. While at the other extreme, organizations carry out their entire business through their website, have adopted a web-based architecture and totally integrated the Internet into all their important business processes.

It is necessary for practising managers to understand the factors responsible for the above differences observed across organizations. Often organizations cannot use the Internet as planned or in an effective manner for their business. An understanding of the factors behind Internet adoption will enable them to evaluate the suitability of their organizational environments and their management strategies to the desired levels of Internet adoption. It will also enable them to plan for and implement Internet adoption in a more effective manner.

While the currently available research studies do describe some factors which influence the phenomenon of Internet adoption by an organization, they do not discuss the collective influence of the factors on the phenomenon. The studies do not also explain the nature of their influence on the level of Internet adoption.

Therefore, these studies do not lead to concrete managerial implications or actionables for managers. At the same time however, rapid acceptance and evolution of the use of the Internet has made it imperative for organizations to utilize it very effectively in their businesses. For effective deployment, they are required to have an understanding of the factors behind successful adoption of the Internet at different levels.

Against this background, this paper intends to:

1. Summarize in a critical fashion the recent literature on the phenomenon of e-business adoption, with focus on the different individual influencing factors.
2. Identify the gaps in the current understanding and opportunities for further research.

2. EXISTING AREAS OF RESEARCH

Examining the existing literature, it is observed that there are numerous strategies adopted by traditional firms to implement operations on the Internet. The decision to adopt a particular strategy was influenced by several factors. We study and analyze those factors in the following sections.

2.1 Internet: Perceived benefits

Internet is now the most popular communication medium for conducting business. Characteristics of the Internet [Eder and Igarria 2001] include:

¾ global, open information and communication platform

¾ not owned by any single organization

¾ flexible and powerful platform for communicating and interactively sharing information

¾ The cost, time and expertise required to connect to this vast network of information is very low

[Beatty et al 2001], [Lal 2002], [Amit and Zott 2001], suggest that firms push for developing their own websites with the intention of realizing potential

benefits like reduced transaction, advertising and distribution costs, elimination of intermediaries, reduced time to complete transactions, capture of market intelligence, faster information to stakeholders and faster responses to new market opportunities. A few of the other expected benefits include mass customization, building of stronger business relationships, greater degree of channel coordination, improved communication and enhanced customer service [Griffith, Plamer 1999], [McKinney 1999].

Another important driver behind Internet adoption has been the increasing trend towards thin-client computing. Thin clients are essentially slimmed-down desktop computers running applications from a central server [Whittle, Sally 2004]. The evolution of the Web has given users an attractive alternative to implement this architecture. As companies move toward Web-enabling internal applications and as enterprise software vendors increasingly provide Web-enabled software, users can access business applications through Web portals and virtual private networks [Mitchell, Robert 2004]. [Costanzo and Chris 2004] discuss the problems with fat client architecture. One of the solutions is adoption of thin-client architecture. Thin client architecture is more scalable, manageable and flexible than a PC-based architecture [Costanzo and Chris 2004].

2.2 Levels of Penetration of the Internet

[Sohn and Wang 1999] talk about how firms implement Internet market / e-Commerce (EC) at different levels, namely, non-adopters, firms planning to adopt, firms with limited users and firms with sophisticated implementation. [Cockburn and Wilson 1995] studied websites of a sample of 500 companies and classified them based on their use as, website providing basic company information, web presence with company information and information about products and services, website carrying company information, products and services information with price details, further step was adding online ordering to above, and website with online ordering and payment. [Liu et al 1997] have in their study identified ways in which large US firms use their corporate web sites to support pre-sales, sales and after-sales business activities. [Dave and Sambamurthy 1999] in their paper discuss how companies have used the WWW platform to capture the capabilities and business potential of this technological innovation.

[Young and Benamati 2000] visited web sites of all fortune 500 countries to identify predominant types of site content and suggested four categories – customer sales, customer service, communication and informational. [Beatty et al 2001] classify organizations into five categories based on the point in time when they adopted the innovation (Internet) relative to other organizations in their social system, namely, pioneers, early adopters, early majority, late majority and laggards.

Another study conducted for Singapore firms, [Timothy et al 2001] proposed four levels of website development – information and catalogues, database, basic transactions, search and integrated site. [Coleman 1998] developed a framework to represent website evolution from the first generation of organization-centric and intranet based homepages to the fifth generation's business-centric and marketplace integration type of delivery platform expansion. [Enders and Jelassi 2000], [Lal 2002] discuss different modes of e-business (EB) transaction, offline (through emails and messaging systems, online (through dynamic URLs), and using shared or individual portals. [Phan 2002] classified EB applications into three categories of electronic markets, inter-organizational systems and customer support applications. In a different study where New Zealand firms were surveyed [Stephan 2001] suggested that primary motivations for EB initiatives and building websites are to list products and services, provide company data, carry out promotional activities, link to alliance partners, and receive customer orders.

[Teo et al 2003] examined factors influencing Internet adoption at four levels – email adoption, web presence, prospecting, business integration and business transformation. [Mahajan et al 2003] studied the intensity of EB adoption in firms and explained the phenomenon using a matrix, along two dimensions, viz., the implemented process and degree of implementation. A different classification of EB information systems based on information integration and timeliness of information was developed by [Pant and Ravichandran 2000]. [Dutta and Segev 1999] examined and classified web sites based on specific features that had been transformed to the Net.

2.3 Adoption of Technology

Many researchers have proposed technological diffusion models explaining causal factors and outcome variables. [Dewar and Dutton 1986] concentrate on the adoption of technological innovation based on parameters like external exposure, complexity, knowledge depth, management attitudes towards change and centralization. [Iacovou and Benbasat 1995], [Crook and Kumar 1998], [Thong 1999], [Premkumar and Ramamurthy 1997] in their papers suggest that EDI adoption is influenced by perceived benefits, organizational readiness and external pressure. [Teo et al 1996] also attempted to identify a set of factors that a firm requires to concentrate on to enhance its ability to employ IT strategically, the factors being IT drivers, business needs, competitor position, external environment, economies of scale and top management guidance. [Kaun and Chau 2001], [Chau et al 1997] conducted further studies to contribute another EDI adoption model using the technology – organization - environment (TOE) framework.

2.4 Factors driving Electronic business adoption by organizations

Various studies have identified factors, which influence the adoption of EB. These factors can be classified into Internal and External factors. Factors internal to the firm are control variables and are within the firm's control. External factors include the firm's external environment and cannot be directly under the control of the organization. These are disturbance variables.

The different factors have been described below.

Internal Factors

ORGANIZATIONAL FACTORS

Studies in this category have identified sub-categories that influence the adoption of EB.

(a) *Inclination towards New Technology.* A favorable attitude towards systems innovation and inclination towards new technology increases the adoption of EC technologies [Iacovou and Benbasat 1995, Sohn and Wang 1999, Mehrtens et al 2001]. [Premkumar et al 1994, 1995] and [Chircu and Kauffman 2000] argue that different organizations feel the need for EC to different extents. Moreover, early adopters of web technologies laid more stress on the technology being consistent and compatible with organizational norms, existing technology, past experiences and present needs than late adopters [Beatty et al 1999, 2001]. In related findings, by [Grover 1993, Premkumar and Ramamurthy 1995], [Crook and Kumar 1998] and [Sohn and Wang 1999] it has been observed that active support of the top management and users lead to EC systems adoption.

(b) *EB Know-how.* EB know-how (executives' knowledge of EB) and Internet skills are complementary to physical assets and provide the business and managerial skills to adopt and manage EB effectively [Zhu and Kraemer 2003]. Organizational learning ability [Mahajan et al 2003] is an important determinant of EB adoption. Firms with superior IT capabilities, technical knowledge and skills [Mehrtens et al 2000], digital literacy [Slywotzky 2000] and IT sophistication [Chwelos et al 2001], [Crook and Kumar 1998] facilitate the adoption of different EC technologies. Existence of champions and better absorptive capacity of the firm facilitates faster and better adoption of new technologies [Sohn and Wang 1999]. [Chircu and Kauffman 2000] suggest that inability to acquire skill and expertise in new technologies, lack of training and education, and inability to learn new technology, form significant barriers to the adoption of EC systems. Organizations with an aggressive and forward-looking technology strategy [Thomson et al 2003] are more likely to leverage new technologies and Internet to respond to market and environment changes.

(c) *Information Intensity.* Organizations with information intensive products will utilize the Internet technology differently than hard goods manufacturers [Palmer and Griffith 1998], [Takacs and Freiden 1998]. The information content of the product and the

information processing content of the production process determine the applicability of Internet. Organizations with information-intensive value chains or value systems will also differ in their adoption of EB than firms with less information-intensive value chains. Product characteristics, the way customers access it and use it, and the information processing required to be carried out by the customer for the same, affect the adoption of EB technologies [Tarafdar and Vaidya 2002]. The requirement of accompanying information for the product, inherent characteristics determining whether it is digitizable or not and information intensity of the product and processes also act as influencing factors. These are findings from studies by [Porter and Millar 1985 and Grover 1993], [Malone et al 1987][Straub and Watson 1995, Choudhury et al 1998], [Palmer and Griffith 1998].

(d) *Organizational readiness.* IT infusion in an organization is influenced by the MIS climate in the organization and MIS maturity [Eder and Igarria 2001]. [Timothy et al 2001], [McIvor 2000] observed that existing capabilities and prior competencies (existing processes and routines, technology, experience with customer groups) make the transition to a new technology easier. Firms with managerial and IT competencies show a strong commitment to EB and website development. [Kuon and Zmud 1987] discussed the importance of internal technology resource for successful IS adoption. Studies that have used the (T-O-E) framework, show that organizations, which are willing to allocate sufficient resources for the development of IS, tend to have a high degree of IS adoption. [Iacovou and Benbasat 1995], [Kuan and Chau 2001] have reported that organizational readiness which is the availability of adequate financial and technical resources, is an important determinant of the extent of IS adoption. The availability of technically skilled human resources is also a crucial factor influencing the adoption of EB [Crook and Kumar 1998, Chircu and Kauffman 2000]. Internal need and top management support [Premkumar and Ramamurthy 1995], pre-existing distribution structure and channel relationship [Palmer and Griffith 1998], determine Internet applicability [Takacs and Freiden 1998], [Gibbs et al 2004].

(e) *Other Organizational Parameters.* A strong grasp of the expected outcomes, issues and requirement of EB facilitates successful adoption of EB. A few authors have reported the effect of various organizational parameters such as business size [Grover 1993], [Hart and Saunders 1997], [Crook and Kumar 1998], [Thong 1999] and centralization [Grover 1993] on the extent of EB adoption. CEO characteristics (innovativeness and IS knowledge), customer loyalty, trust with regard to the integrity of the information and security of transactions conducted on the Internet, [Thong 1999], [Choudhury et al 1998], are also influencing factors. More informed firms, with greater scope [Zhu et al 2003] are more

likely to adopt EB. Cultures of existing business and EB should gel well.

TECHNOLOGY CHARACTERISTICS

The characteristics of the Internet and other new technologies influence organizational decisions for EB adoption. IS characteristics like complexity, relative advantage, compatibility and the perceived benefits also act as facilitators. Additionally, EB technologies which are compatible with existing applications, hardware and software platforms, and with existing cultural value (related to technology) [Thompson 2003] are more likely to be adopted. Different level of adoption need different combinations of Internet related technology. These findings have been reported by [Grover 1993], [Premkumar et al 1994], [Beatty et al 2001], [Thong 1999], [Iacovou 1995], [Kuan and Chau 2001], [Gibbs et al 2004] and [Zhu et al 2004].

PERCEIVED BENEFITS

Organizations pursue EB adoption with the intention of realizing operational and strategic benefits. [Mehrtens et al 2001] have found that organizations realizing the greater information reach offered by websites, actively adopt the Internet as a communication medium. Direct perceived benefits as discussed in the section earlier and expectation of indirect benefits (customer service, productivity increase), also influence the extent to which organizations adopt EB / EC technologies [Premkumar et al 1994, Crook and Kumar 1998, Kaplan and Sawhney 2000, Beatty et al 2001, Chwelos et al 2001, Mehrtens et al 2001].

External Factors

Pressures from the firm's external environment, competitors' moves and customer expectations have been a major factor influencing the adoption of EB.

ORGANIZATIONAL ENVIRONMENT

Competitive pressures drive an organization to adopt EB [Grover 1993], [Iacovou and Benbasat 1995], [Premkumar and Ramamurthy 1995], [Sohn and Wang 1999] and [Teo et al 2003]. Support and motivation from customers and suppliers, their changing orientation also help organizations to move to EB systems [Grover 1993], [Iacovou and Kumar 1995], [Crook and Kumar 1998], [Hart and Saunders 1998], [Chwelos et al 2001], [Mehrtens et al 2001]. A higher level of consumer readiness (customer willingness and Internet penetration) drives firms to adopt EB [Zhu et al 2003]. Lack of trading partner readiness to engage in Internet transactions is a significant inhibitor for EB adoption. EB intensity of the country [Zhu et al 2003] and industry level [Young and Benamati 2000] factors also act as drivers. [Zhu et al 2004], [Gibbs et al 2004] find legislation barriers, external pressure, government promotion, policy environment and country effects as significant predictors of scope of use of EB technology. [Hart and Saunders 1997, 1998] observe that trust and interdependence between a company and its suppliers

and customers are facilitators for EC adoption.

3. INTERNET ADOPTION BY ORGANIZATIONS: A FRAMEWORK

As we have seen, various classifications of websites have been proposed in the literature. But there is no direct explanation for the differences between them. Most other studies just discuss the intent to adopt EB and the influencing drivers.

Figure 1. below represents a possible framework explaining the phenomenon of Internet adoption. This framework emerges out of our observations.

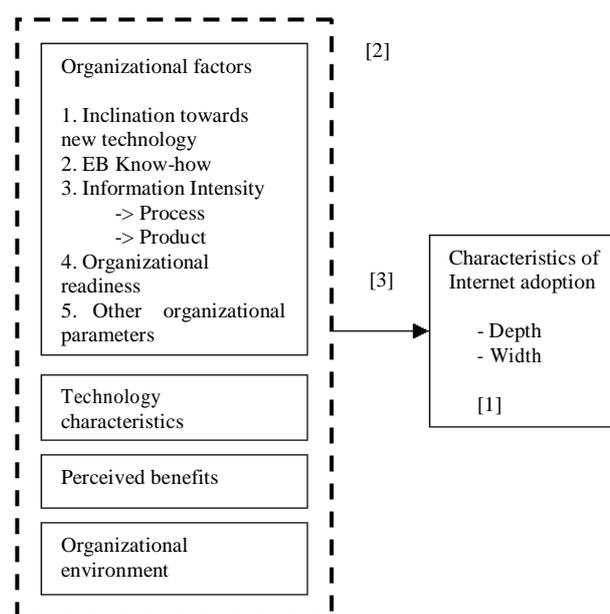


Figure 1. Future research directions

The literature emphasizes the differences in activities that organizations have implemented on the Internet. However, it does not take explicit cognizance of the differences in the involvement of the different stakeholders in the Internet-based business (Refer [1] in Fig.1).

Also, while the literature identifies the individual causal factors in a rather exhaustive manner, it does not explicitly address the nature of their influence on the phenomenon of Internet adoption, their collective influence and the interactions among themselves (Refer [2] in Fig.1).

Information Intensity, particularly along the value system, is a very important determinant of EB adoption but has not been discussed much in EB adoption models (Refer [2] in Fig.1).

Finally, the literature does, very justifiably, try to classify websites, which represents at least partially the degree of Internet adoption. However, the

conceptual linkages between the relevant drivers and the different characteristics of the websites have not been studied in depth (Refer [3] in Fig.1).

4. CONCLUSION

It is observed that there are significant differences in the level of Internet adoption among organizations. Also, often organizations cannot use the Internet as planned or in an effective manner for their business. In this paper we have tried to examine the factors responsible for successful EB adoption and identified gaps in the present literature. Subsequently we have developed a possible framework for explaining the phenomenon of Internet adoption, which also suggests directions for future research efforts. Such analysis would provide organizations an adequate understanding of the factors that would enable them to evaluate the suitability of their organizational environments and their management strategies to their desired levels of Internet adoption.

REFERENCES

- [1] Beatty, Shim, Jones, "Factors influencing corporate web site adoption: a time based assessment", *Information & Management*, (38), pp. 337-354, 2001
- [2] Chatterjee, Debabroto (Dave), Sambammurthy, V., "Business Implications of Web Technology: An Insight into Usage of the WWW by U.S. Companies", *Electronic Markets*, (9:2), 1999
- [3] Chau, P., Y., K., Tam, K., Y., "Factors Affecting the Adoption of Open Systems: An Exploratory Study", *MIS Quarterly*, March 1997, pp.1-24
- [4] Chircu, A. M., and Kauffman, R. J., "Trust, Expertise and EC intermediary adoption", *Proceedings of the 2000 Americas Conference on Information Systems*
- [5] Chircu, A. M., and Kauffman, R. J., "Limits to Value in Electronic Commerce-Related IT Investments", *Journal of Management Information Systems* (17:2), Fall 2000, pp. 59-80.
- [6] Choudhury, V., Hartzel, K., and Konsynski, B. R., "Uses and Consequences of Electronic Markets", *MIS Quarterly*, December 1998, pp. 471-507.
- [7] Chwelos, P., Benbasat, I., and Dexter, A., "Research Report: Empirical Test of an EDI Adoption Model", *Information Systems Research* (12:3), pp. 304-321, 2001
- [8] Cockburn C., Wilson T.D., "Business Use of the World-Wide Web", *Information Research*, (1:2), 1995
- [9] Coleman K., "Make your website a business success", *EB Advisor*, pp.12-17, 1998
- [10] Costanzo, Chris, "Thin-Client Strategy Fuels Ky. Bank's Growth", *American Banker*, (169:119), 2004
- [11] Crook, C. W., and Kumar, R. L. "Electronic data interchange: A multidisciplinary investigation using grounded theory", *Information and Management* (34), 1998, pp. 75-89.
- [12] Dewar R., Dutton J., "The adoption of radical and incremental innovations: an empirical analysis", *Management Science*, 1986
- [13] Dien D. Phan, "EB success at Intel: an organization ecology and resource dependence perspective", *Industrial Management and Data Systems*, Vol. 102, No.4, 2002, pp. 203-210
- [14] Dutta, S., Arie Segev, "Business Transformation on the Internet", *European Management Journal*, Volume: 17, Issue: 5, October, 1999
- [15] Enders, A., Jelassi, T., "The Converging Business Models Of Internet and Brick-and-Mortar Retailers", *European Management Journal*, (18:5), 2000, pp.542-550
- [16] Eder, Lauren & Igbaria, Magid, "Determinants of Intranet Diffusion and Infusion", *Omega*, June 2001, Volume 29, No. 3, pp. 233-242.
- [17] Fang, Wu, Mahajan, V., Balasubramaniam, S., "An Analysis of EB Adoption and Its Impact on Business Performance", *Journal of the Academy of Marketing Science*, 2003, (31:4), pp. 425-447
- [18] Gibbs, Jennifer, Kraemer, Kenneth L., "A Cross-Country Investigation of the Determinants of Scope of EC Use: An Institutional Approach", *Electronic Markets*, Vol. 14, No., 2004
- [19] Griffith, David A., Plamer, J.W., "Leveraging the Web for Corporate success", *Business Horizons*, Jan-Feb 1999, pp. 3-10
- [20] Griffith, David A., Plamer, J.W., "Information Intensity: A Paradigm For Understanding Web Site Design", *Journal of Marketing THEORY AND PRACTISE*, Summer 1998, pp. 38-42
- [21] Grover, V., "An Empirically derived Model for the Adoption of Customer-based Inter-organizational Systems", *Decision Sciences* (24:3), 1993, pp. 603-638.
- [22] Hart, P., and Saunders, C. "Emerging Electronic Partnerships: Antecedents and Dimensions of EDI Use from the Supplier's Perspective", *Journal of Management Information Systems* (14:4), 1998, pp. 87-111.
- [23] Hart, P., and Saunders, C. "Power and Trust: Critical Factors in the Adoption and Use of Electronic Data Interchange", *Organization Science* (8:1), 1997, pp. 23-42.
- [24] Iacovou, C. L., Benbasat, I., and Dexter, A. S. "Electronic Data Interchange and Small Organizations: Adoption and Impact of Technology", *MIS Quarterly*, December 1995, pp. 465-485.
- [25] Kaplan, S., and Sawhney, M. "E-Hubs: The New B2B Marketplaces", *Harvard Business Review*, May-June 2000, pp. 97-103.
- [26] Kuan, K., Chau, P., "A Perception-based Model for EDI adoption in small businesses using a technology-organization-environment framework", *Information & Management*, 2001, 38, pp.507-521
- [27] Kuan T. H., Zmud R., "Unifying the fragmented Models of Information Systems Implementation", *Critical Issues in IS Research*, John Wiley, NY, 1987
- [28] Kowtha, N., R., Choon, Timothy, "Determinants of Website Development: A Study of Electronic Commerce in Singapore", *Information & Management*, 2001, 39, pp. 227-242
- [29] King, W., R., Teo, S., H., Thompson, "Key

- Dimensions of Facilitators and Inhibitors for the Strategic Use of Information Technology”, *Journal of Management Information Systems*, Spring 1996, (12:4), pp.35-53
- [30]Lal K., “EB and Export Behavior Evidence from Indian Firms”, Discussion Paper No. 2002/68, 2002
- [31]Liu, C., Arnett, K., Capella, Beatty, R., “Web sites of the Fortune 500 companies: Facing customers through home pages”, *Information & Management*, 1997
- [32]McIvor, R., “Electronic commerce: re-engineering the buyer supplier interface”, *Business Process Management Journal*, Vol. 6, 2, 2000
- [33]Malone, T. W., Yates, J., and Benjamin, R. I., “Electronic Markets and Electronic Hierarchies”, *Communications of the ACM* (30:6), June 1987, pp. 484-497
- [34]Mitchell, Robert L., “Through Thick and Thin”, *Computerworld*, Vol. 38, Issue 30
- [35]Mehrtens, J., Cragg, P. B., and Mills, A. M. “A Model of Internet Adoption by SME’s”, *Information and Management* (39), 2001, pp. 165-176.
- [36]Michael E. Porter, “Strategy and the Internet”, *Harvard Business Review*, 2001
- [37]Porter, M. E., and Millar, V. E., “How information gives you competitive advantage”, *Harvard Business Review*, 1995, pp. 149-160.
- [38]Premkumar, Ramamurthy, Crum, “Determinants of EDI adoption in the transportation industry”, *European Journal of Information Systems*, 1997
- [39]Premkumar, G., Ramamurthy, K., and Nilakanta, S. “Implementation of Electronic Data Interchange: An Innovation Diffusion Perspective”, *Journal of Management Information Systems* (11:2), 1994, pp 11-23.
- [40]Premkumar, G., and Ramamurthy, K. “The role of inter-organizational and organizational factors on the decision mode for adoption of inter-organizational systems”, *Decision Sciences* (26:3), 1995, pp. 303-337.
- [41]Riemenschieder, R., Mckinney, V., “Assessing the Adoption of Web-Based EC for Businesses: A Research Proposal and Preliminary findings”, *Electronic Markets*, 1999, Volume 9:1/2, pp. 9-13
- [42]Sean Xu, Zhu K., Gibbs J., ”Global Technology, Local Adoption: A Cross-Country Investigation of Internet Adoption by Companies in the United States and China”, *Electronic Markets*, 2004, vol. 14, no. 1, pp. 13-24
- [43]Sohn C., T.W. Wang, “Diffusion factors and adoption levels of the Internet market: empirical analysis for the computer retailing companies”, in: *Proceedings of the 1999 Americas Conference on Information Systems*, pp. 259-261.
- [44]Slywotzky, A. J., “The Age of the Choiceboard: in The Future of Commerce”, *Harvard Business Review*, January-February 2000.
- [45]Stephen drew, “EB Research practice: towards an agenda”, *Electronic Journal of Business Research Methods*, 2002
- [46]Straub, D. W., and Watson, R. T., “Research Commentary: Transformational Issues in Researching IS and Net-Enabled Organizations”, *Information Systems Research* (12:4), December 2001, pp. 337-345.
- [47]Somendra Pant, T. Ravichandran, “A framework for planning EB Information Systems: A Research proposal”, *Americas Conference on Information Systems* 2000
- [48]Stephen B., Clark D., Corner P., Gibb J., Kate K., and Pavlovich K., “Adoption and Implementation of EB in New Zealand: Preliminary Results”, *Proceedings of the 9th Annual Conference of the New Zealand Strategic Management Society*, 2001 pp.129-136
- [49]Tarafdard, M., and Vaidya, S. D., “Research in Organizational Adoption of EC- The Need for an Interdisciplinary Perspective”, *Proceedings of the American Conference on Information Systems*, 2002
- [50]Takacs, S., Freiden, J., “Changes on the Electronic frontier: Growth and opportunity of the World Wide Web”, *Journal of Marketing Theory and Practice*, Summer 1998, pp. 24-38
- [51]Teo, Thompson SH, Pian, Yujun, “A contingency perspective on Internet adoption and competitive advantage”, *European Journal of Information Systems* (2003) 12, pp. 78-92
- [52]Thong, J., “An Integrated Model of Information Systems Adoption in Small Businesses”, *Journal of Management Information Systems*, Spring 1999, (15:4), pp. 187-214
- [53]Whittle, Sally, “Think thin, think efficient”, *IT Training*, Jul/Aug2004
- [54] Young D., & Benamati J., “Differences in Public Websites: the current state of large US firms”, *Journal of E-commerce research*, 2000(3)
- [55]Zhu, K., Kraemer, K., Xu, S., “Electronic business Adoption by European firms: A cross country assessment of the facilitators and inhibitors”, *European Journal of Information Systems* (2003) 12, pp.251-268
- [56]Zott, C., Amit, R., Donlevy, J., “Strategies for Value Creation in E-Commerce: Best Practise in Europe”, *European Management Journal*, 2000, (18:5), pp. 463-475

Towards a Definition of IT Management Sophistication in Small Firms

Theekshana Suraweera, Paul Cragg, Annette Mills

Department of Accountancy, Finance and Information Systems
University of Canterbury, Christchurch, New Zealand
theek.suraweera@canterbury.ac.nz

ABSTRACT

This paper reports the results of a study aimed at understanding and characterising IT management sophistication in small businesses. A multiple-case study approach was therefore used to explore the IT management practices of small businesses. Data collection involved four small-medium chartered accountancy (CA) firms and three IT firms closely associated with the CA industry. The case results combined with past research suggested 30 individual indicators of IT management sophistication. Further investigation of these 30 indicators yielded five underlying factors characterising IT management sophistication, namely *planning*, *controlling*, *organising*, *leading* and *external expertise*. The results provided a useful foundation for examining the impact of IT management sophistication on organisational performance and developing appropriate improvement strategies. Directions for future research are also discussed.

Keywords: small businesses, IT management, IT management sophistication, accounting firms

1. INTRODUCTION

The significance of the small business sector in national economies has been well documented (e.g. Vos & Nyamori, 1997). Computer based information systems have grown in importance to small firms (Fuller, 1996), and are now being used increasingly to help them compete (Mata, Fuerst, & Barney 1995). IS researchers have also emphasised the importance of managing the IT resource in small businesses in an effective manner (Raymond & Pare, 1992; El Louadi, 1998). Although the significance of IT management in the context of small businesses has been emphasised in the IS literature, so far a consensus has not been reached with respect to the definition of this construct. Further, a valid instrument to measure the 'IT management sophistication in small firms' construct is yet to be developed.

This study aimed to characterise 'IT management' in the context of small firms with a view to exploring the concept of 'IT management sophistication', that is, of some firms being more sophisticated than others in their approach to IT management. This objective was addressed using case study research of four small chartered accountancy firms in New Zealand. The case studies explored the practical and operational issues related to IT management in small businesses, to derive a pool of indicators that can be used to measure IT management sophistication. This paper reports the results of the case studies, suggesting a pool of operational indicators that can be used to characterise the construct of IT management sophistication in small business.

This paper is organised under five sections. The literature review is presented next and the research approach is outlined in section 3. This is followed by a discussion of the case study data and the resultant pool of indicators that characterise IT management

sophistication in small businesses. Research conclusions and limitations are presented in section 5.

2. LITERATURE REVIEW

Table [1] summarises the existing frameworks that suggest key aspects of IT management. These frameworks provide useful insights for identifying potential areas and sub-functions of IT management that could be adapted for this study.

The diversity of IT management dimensions identified in Table [1] shows that there is no commonly accepted definition of the term 'IT management'. Planning, organising and controlling are common to many of the frameworks characterising IT management, although some include aspects not shared by the others. For example Feeny & Willcocks (1998) identify IS/IT leadership, relationship building and business systems thinking as core IS capabilities, but not Earl (1989), Nolan (1973) and Guptha et al., (1997). Furthermore, while large firms are concerned with IT management issues such as architecture planning, contract monitoring and IS function management (Feeny & Willcocks, 1998), small firms deal with issues such as educating the users, involvement of external consultants and implications of top management (Pollard & Hayne, 1998). The importance of external expertise in the management and implementation of information systems, especially in the small business sector, has been highlighted by several researchers (e.g. Fink, 1998; Gable, 1996; Thong et al., 1996).

Table [1] Key Aspects of IT Management

Earl (1989)	Nolan (1973)	Boynton et al. (1994)	Galliers & Sutherland (1991)	Davis et al. (1995)
<ul style="list-style-type: none"> • Planning • Organising • Control 	<ul style="list-style-type: none"> • Planning • Organising • Control 	<ul style="list-style-type: none"> • Project management • Strategic Management • Service Control • Service Planning • Resource Planning • IS services • IS Function Management • Development Maintenance 	<ul style="list-style-type: none"> • Strategy • Structure • Systems • Staff • Style • Skills • Super-ordinate goals 	<ul style="list-style-type: none"> • IS and business alignment • IS staff skills • Competitive advantage • End-user computing • Telecommunications • Security and control • Identifying projects • Measuring IS effectiveness • Open systems • Application proliferation
Feeny & Willcocks (1998)	Sabherwal & Kris (1994)	Guptha et al. (1997)	Pollard & Hayne (1998)	Raymond & Pare (1992)
<ul style="list-style-type: none"> • IS/IT Leadership • Business systems thinking • Relationship building • Architecture planning • Making technology work • Informed buying • Contract facilitation • Contract monitoring • Vendor development 	<ul style="list-style-type: none"> • IS planning • Top management involvement (in planning) • IS performance evaluation • IS manager's knowledge of business plans • Top management's knowledge of IT 	<ul style="list-style-type: none"> • IT planning mode • IT control mode • IT organisation • IT integration 	<ul style="list-style-type: none"> • IS for competitive advantage • IS project management • Software development • Responsive IT infrastructure • Aligning IS • Technological change • Communication networks • Business process redesign • Educating users • IS human resource 	<ul style="list-style-type: none"> • Organisational objectives • Top management implications • IT investment • IT adoption • Presence of consultants • IT planning • IT control • IT evaluation • IS personnel • Role of IS function • Decision level • Type of development • Position of IS

Although Fink (1998) asserted that the management effort towards IT in small firms is negligible compared to large firms, it is not proper to infer that small businesses have absolutely no practices to manage their IT. For example, Rodwel & Shadur (1997) confirmed that the practices related to human resources management are more sophisticated than one may be led to believe. However, Cragg & King (1993) found that with the maturity of small firms, only minor changes had been observed in the management of IT while many small firms had experienced growth with respect to the number and type of IT applications.

Raymond & Pare (1992) asserted that IT sophistication may be characterised under two major dimensions: IT usage and IT management (See Figure [1]). This paper focuses on the IT management dimensions to better understand the concept of IT management as it applies to small firms. Although there are certain commonalities in the characterisations of IT management, it is evident that a commonly accepted means to define IT management in small firms has not evolved. Examination of two small firm studies, that is, Pollard & Hayne (1988) and Raymond & Pare (1992) cited in

Table [1] showed little agreement with respect to major issues contributing to the characterisation of IT management.

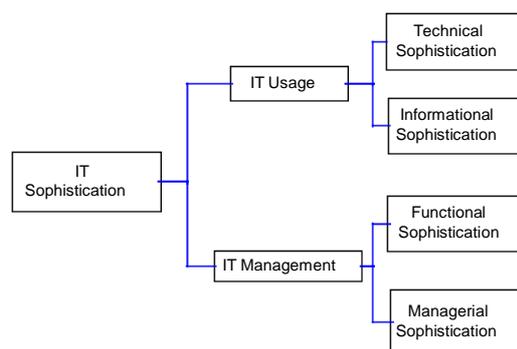


Figure [1] Dimensions of IT sophistication (Raymond & Pare, 1992, p. 7)

Table [1] also suggests that the general management sub-themes, *planning, organising, controlling and leading* (Daft, 1988, Earl, 1989, Nolan, 1973, Schermerhorn, 1989) with respect to IT management

may provide a useful basis for characterisation of this concept in broad terms.

3. RESEARCH APPROACH

As indicated above, this research relied on the case study method. This method found to be useful in small firm studies and also in IS research leading to theory development (Ramano, 1989, Yin, 1994, Zinatelli & Cavaye, 1994). Case studies were conducted in four small and medium sized chartered accountancy (CA) firms in Christchurch, New Zealand.

Chartered accounting firms offer accounting and other services to individuals and other firms. For example, many CA firms help other small firms create their annual accounts. They also offer business advice, for example, on the structure of a firm, tax liability, and employment issues. The intensive use of IT by CA firms for information processing in their main business activity and in producing information for clients, and the similarity of technology used across the firms in the industry, were the major reasons for choosing CA firms for this study.

Case study interviews were conducted between February to May 1999. Number of partners in this sample of firms ranged from 1 to 9 and the full-time accounting staff ranged from 3 to 50. Two of these firms were using APS accounting software, and one used CA systems (MYOB Office Systems). Three of these firms employed external IT consultants. (The case study firms are identified as CAF1, CAF2, CAF3 and CAF4 in this paper).

The case studies were aimed at identifying the issues representing the key IT management practices in small firms to explore the concept of IT management. The information gathered from the case studies was supplemented with comprehensive discussions with two IT consultants, who were closely associated with CA firms, and the general manager of a major software supplier to the chartered accountancy industry in New Zealand and Australia. (They are identified as CON1, CON2 and CON3 in this paper).

4. DATA ANALYSIS AND DISCUSSION

The case study data comprised mainly of transcribed interviews. There were two major phases to the case study data analysis process. Firstly, the case study data were scrutinised to identify the issues, concerns and important points that were indicators of IT management practices in small firms. The preliminary pool of likely indicators so derived was examined in the light of past research findings. Consideration was given in this review process to examining if a particular issue was actually IT management related. Emphasis was placed on those issues which were identified in more than one case and which have also been dealt with by past

researchers. Also, a few items were added to this pool of indicators to reflect certain critical IT management issues that were strongly supported by past researchers, although not brought to light through the case studies. Table [2] lists four draft item statements derived in this phase of the analysis. (Due to page restrictions, the full set of 30 items is not provided here but is available from the first author).

The second phase of the case study analysis focused on assessing the content validity and to determine the potential sub-dimensions of IT management sophistication represented by the 30 indicator statements. Similar to the approach adopted by Chan et al. (1992), the item statements were reviewed by nine experts in this phase of the study. These experts represented academics/researchers and technical staff from the IT/IS and Management disciplines. In the review process each expert was provided with a set of cards on which one item statement was printed. They were asked to sort the cards into a number of piles representing the likely dimensions of IT management. The four major sub-functions of management (i.e. planning, organising, controlling and leading) were suggested as a basis for investigation. However, the experts were encouraged to present their own views about the factors that may represent IT management sophistication. The sorting exercise was followed by detail discussions with the experts.

All participants of the expert opinion review agreed that these 30 statements, in general, would reflect the important issues of IT management in small firms. They also agreed that examination of these issues would provide a sufficient understanding of IT management in small firms, which in turn could be used to measure the level of IT management sophistication.

Participants generally agreed that the concepts underlining the statements presented can be categorised into the four broader factors of planning, controlling, organising and leading, and most statements were classified under the four factors accordingly. However, the participants were not always unanimous in their classification decisions with respect to certain statements. For example, the item "*We use a rigorous IT planning process within our firm*" was classified under planning by seven participants; whereas one participant placed this item under controlling and another under organising. Such issues were resolved by reviewing them in the light of past research references (e.g. definitions of planning, organising, controlling and leading by Schermerhorn, 1989).

At the end of the review process, there were three items that did not appear to fit into any of the four factors identified above. All three item statements were related to obtaining support from external expertise and gathering information from parties external to the firm. Presence

Table [2] Draft Item Statements, Case Study Evidence and References (a Sample)

<p>Item #1</p> <ul style="list-style-type: none"> • We have comprehensive procedures for maintaining the security of information stored on the computer. (<i>Davis et al., 1995 and Raymond et al., 1992</i>)
<ul style="list-style-type: none"> • It's amazing how much faith you're putting in your staff in small businesses and how little security; ... [of] the most valuable part of the business; the data you have... (CON2) • [A] lot of firms are not backing-up [their data]. Quite recently we had a major problem with our product. ... [Due to some unforeseen problem] the data were corrupted [when the new system was up and running for the first time]. About twenty firms around the country got affected by it. ... and still believe that a lot of firms have very very poor security. (CON3) • At the moment internally we don't really have much [internal] security. (CAF4) • We're trying to convince most people to take care of their data. The importance of security of data. (CON1)
<p>Item #2</p> <ul style="list-style-type: none"> • We have comprehensive procedures for managing the use of IT resources. (<i>Raymond et al., 1992</i>)
<ul style="list-style-type: none"> • No, there's no procedure written down; it's really just casual knowledge.(CAF3) • ..no we don't write too much [procedures] down.. Because by the time you write it down, it would have changed.. Personally I don't believing in writing things down too much. (CAF4) • There is no control over what software has been loaded on the PCs... ; there is often different versions of all software on all the PCs depending on when they brought them. (CON2)
<p>Item # 3 and # 4:</p> <ul style="list-style-type: none"> • We closely monitor the progress of our IT projects. • We closely monitor the performance of the IT function. <p>(<i>Raymond et al., 1992; Sabherwal & Kris, 1994; Boynton et al. 1994; Pollard & Hayne., 1998; Feeny & Willcocks., 1998; Gupta et al., 1997</i>)</p>
<ul style="list-style-type: none"> • We've got meetings with all the staff every fortnightProbably our monitoring is based on identifying how many problems we have. (CAF1) • We monitor obviously the profitability of the firm and we could tell that really within six months [whether] we have actually recovered [the cost of a particular IT investment] (CAF1). • Probably in three months down the track they [CA firms/software users] should ask for some consultation and review from the supplier. (CON3)

of IT consultants (Raymond et al., 1992) and obtaining advice from external experts (Thong et al., 1996, Yap et al., 1992, yap et al., 1994) have been identified as significant feature in IT management in small business. Cragg & King (1993) also identified the consultant's support as a factor that encourage IT growth. Referring back to the case studies, it could be seen that the CA firms gather external information from other means such as involving in the software user groups, mostly for purposes related to managing IT, and not for other operational activities of CA firms. Research shows that SMEs also value information about IT provided by external sources (e.g. vendors and consultants) for successful IT adoption (Fink 1998). Fann & Smeltzer (1989) have shown that suppliers/vendors and competitors are important sources of information for operational decision making in small firms. Accordingly, a fifth factor, namely *external expertise*, was proposed for characterising IT management sophistication in small business. The finalised list of item statements classified under the likely major factors (i.e. sub-dimensions) of IT management sophistication in small firms is given in Table [3].

5. CONCLUSIONS

This case study research examined IT management

within small businesses. The study identified 30 item statements/indicators of IT management sophistication. These 30 items were then classified into five factors: planning, organising, controlling, leading, and external expertise. These results fit well with the findings from past research. Many of the indicators relating to planning, controlling and organising have been used in prior research to characterise IT management sophistication. Leading emerged as a relevant factor of IT management sophistication in small business according to the current study. Past research and the case study evidence also supported the inclusion of indicators pertaining to *external expertise*.

The development of 30 items statements representing, and identifying the associated five major factors/sub-dimensions are the major contributions of this research. IS researchers can use them to characterise the 'IT management sophistication in small business' construct and further explore the relationships between this and other related constructs, such as IS success, IT enabled organisational performance and competitive advantage. Practitioners could use the results to help them determine strengths and weaknesses of IT management in small firms, and then formulate appropriate strategies aiming at achieving organisational success and to gain competitive advantage.

A major limitation of this research was the restriction of the study to four small and medium sized chartered accountancy firms in New Zealand. Given the diversity within the small business sector, this element restricts the generalisation of findings. Further investigation, by way of additional case studies, may be directed towards clarifying the role of external expertise in IT management sophistication in small business. Ascertaining the validity of the derived item statements as measures IT management sophistication in small business construct is another direction for further research. Survey research using the derived measures

(the 30 item statements in table [3]) of IT management sophistication in small firms and data analysis aiming at instrument development and validation are suggested to peruse this end.

ACKNOWLEDGEMENT

This research was supported by the department of Accountancy, Finance and Information Systems of the University of Canterbury, New Zealand.

Table [3] Factors and Indicators of IT Management Sophistication in Small Business

Factors /Item Statements
<p>IT Planning</p> <ul style="list-style-type: none"> • Our firm is continuously searching for and evaluating new IT developments for their potential use in the firm • We use a rigorous IT planning process within our firm. • Mostly, our IT plans are written. • Our IT plans are very detailed. • Our IT plans are frequently reviewed to accommodate the changing needs of the firm. • Our firm recognises IT planning as an important part of the overall business planning process. • Our IT system is designed to be closely aligned with the overall objectives of the firm. • In our firm IT is used to improve the firm's competitive position.
<p>IT Organising</p> <ul style="list-style-type: none"> • We have one or more staff members who spend most of their time managing our firm's IT resources. • We select our IT vendors and external consultants according to formal criteria (<i>eg based on a combination of their proven success, IT expertise, familiarity with our own line of business</i>). • In our firm, staff participates in making major IT decisions. • Our firm has a flexible approach to organising IT operations and maintenance. • We select the most suitable package based on proven success, when it comes to software acquisition.
<p>IT Controlling</p> <ul style="list-style-type: none"> • We closely monitor the performance of our IT systems. • We closely monitor the progress of our IT projects. • We have comprehensive procedures in place for maintaining the security of information stored in our computers. • We have comprehensive procedures in place for controlling the use of IT resources (<i>eg who can use specific software, who has access to specific databases</i>). • In our firm the roles and responsibilities for IT direction and development are clearly defined. • In our firm the roles and responsibilities for IT operations are clearly defined. • We have formal procedures for the acquisition and development of new IT systems.
<p>IT Leading</p> <ul style="list-style-type: none"> • IT management in our firm is characterised by strong leadership. • Our top management plays an active role in addressing the firm's IT issues. • Our top management perceives as IT is critical to our business success. • Our managers have created a vision among the staff for achieving IT objectives. • Our managers have inspired staff commitment towards achieving IT objectives. • Our managers have directed the efforts of staff towards achieving IT objectives. • Our firm is committed to providing staff with appropriate IT training.
<p>External IT Expertise</p> <ul style="list-style-type: none"> • We have very effective working relationships with our IT vendors and/or external consultants. • We gather IT information from others in the industry. • Our firm relies heavily on external IT expertise.

REFERENCES

1. Boynton, A. C., R. W. Zmud et al., The Influence of IT Movement Practice on IT Use in Large Organisations, *MIS Quarterly*, Vol. 18, No. 3, pp299-318, 1994.
2. Chan, Y. E. *Business Strategy, Information Systems Strategy, and Strategic Fit: Measurement and Performance Impacts*, Unpublished PhD dissertation, School of Business Administration. London, University of Western Ontario: 609, 1992.
3. Cragg, P. B., M. King, Small-Firm Computing: Motivators and Inhibitors, *MIS Quarterly* Vol. 17, No. 1, pp47-60, 1993.
4. Daft, R. L., *Management*. New York, The Dryden Press. 1988.
5. Davis, J. D., R. Menon, S R Mundy, B.G.Thomson, Key Issues in Information Systems Management: A New Zealand Perspective, *New Zealand Journal of Computing*, Vol. 6 No. 1A, pp15-21, 1995.
6. Earl, M. J., The Risks of Outsourcing IT, *Sloan Management Review*, Vol. 37, No. 3, pp26-32, 1996.
7. El Louadi, M. E. The Relationship among Organisation Structure, Information Technology and Information Processing in Small Canadian Firms, *Canadian Journal of Administrative Sciences*, Vol. 15, No. 2, pp180-199, 1998.
8. Fann, G. L., L. R. Smeltzer, The Use of Information from and about Competitors in Small Business Management. *Entrepreneurship: Theory and Practice (Summer)*, pp35-46, 1989.
9. Feeny, D. F., L. P. Willcocks, Core IS Capabilities for Exploring Information Technology, *Sloan Management Review*, Vol. 39, No. 3, pp9-21, 1998.
10. Fink, D., Guidelines for the Successful Adoption of Information Technology in Small and Medium Enterprises, *International Journal of Information Management*, Vol. 18, No. 4, pp243-253, 1998.
11. Fuller, T., Fulfilling IT Needs in Small Business; A Recursive Learning Model, *International Small Business Journal* Vol. 14, No. 3 (issue no. 55), pp25-44, 1996.
12. Gable, G. G., Outsourcing of IT Advice: A Success Prediction Model, 1996 Information Systems Conference of New Zealand, Palmerton North, New Zealand, IEEE, 1996.
13. Galliers, R. A. R Sutherland. Information Systems Management and Strategy Formulation: The Stages of Growth Model Revisited, *Journal of Information Systems*, Vol. 1, No. 2, pp 89-114, 1991.
14. Guptha, Y. P., J. Karimi, M. Somers, Alignment of Firms Competitive Strategy and Information Technology Management Sophistication: the Missing Link, *IEEE Transactions of Engineering Management*, Vol. 44, No. 4, pp399-412, 1997.
15. Mata, F. J., W. L. Fuerst, J. B. Barney, Information Technology and Sustained Competitive Advantage: A Resource Based Analysis, *MIS Quarterly*, Vol. 19, No. 5, pp 487-505, 1995.
16. Nolan, R. L., Managing the Computer Resources: A Stage Hypothesis, *Communications of the ACM*, Vol. 16, No. 7, pp399-405, 1973.
17. Pollard, C. E., S. C. Hayne, The Changing Face of Information System Issues in Small Firms, *International Small Business Journal*, Vol. 16, No. 3 (Issue No. 63), pp 71- 87, 1998.
18. Raymond, L., G. Pare, Measurement of Information Technology Sophistication in Small Manufacturing Business, *Information Resources Management Journal*, Vol. 5, No. 2, pp 4-15, 1992.
19. Rodwell, J., M. Shadur, What's Size Got to do with it? Implications for Contemporary Management Practices in IT Companies, *International Small Business Journal*, Vol. 15, No. 2, pp51-62, 1997.
20. Romano, C. A., Research Strategies for Small Business: A Case Study Approach, *International Small Business Journal*, Vol. 7, No. 4, pp 35-43, 1989.
21. Sabherwal, R., P. Kris, The Alignment between Organisational Critical Success Factors and Information Technology Capability in Academic Institutions, *Decision Sciences*, Vol. 25, No. 2, pp301- 330, 1994.
22. Schermerhorn, J. R., *Management for Productivity*, New York : John Wiley & Sons, 1989..
23. Thong, J. Y. L., C. Yap, et al., Top Management Support, External Expertise and Information Systems Implementation in Small Business, *Information Systems Research*, Vol. 7, No. 2, pp248-267,1996..
24. Vos, E., R. O. Nyamori, An Empirical Analysis of the Capital Structure of New Zealand Firms, *Small Enterprise Research*, Vol. 5, No. 2, pp29-38, 1997.
25. Yap, C., C. P. P. Soh, et al., Information System Success Factors in Small Business, *Omega, The International Journal of Management Science*, Vol. 20, No. 5/6, pp597-609, 1992.
26. Yap, C., J. Y. L. Thong, et al., Effect of Government Incentives on Computerisation of Small Business, *European Journal of Information Systems*, Vol. 3, No. 3, pp191-206, 1994.
27. Yin, R. K., *Case Study Research Design and Methods*, California: Sage, 1994.
28. Zinatelli, N., A. Cavaye, Case Study Research in Information Systems: Strategies for Success and Pitfalls to Avoid, *New Zealand Journal of Computing*, Vol. 5, No. 1, pp17- 22, 1994.

Towards a Framework of Enterprise Integration: Linking Quality Management, Organizational Learning and Information Technology

Jiunn Chieh Lee¹

¹ Department of Information Systems and Operations Management, The Auckland Business School, The University of Auckland, Auckland, Private Bag 92019, Auckland 1001, New Zealand
jc.lee@auckland.ac.nz

ABSTRACT

This paper proposes a framework that can be used to explain and improve enterprise integration practices. It draws on the traditions of quality management and organizational learning to understand how implementation of advanced information technologies such as enterprise resource planning and customer relationship management may be explained and improved. Enterprise systems implementations at two subsidiaries of two separate large conglomerates are used to illustrate this framework. In particular, it is posited that quality management and organizational learning principles and practices are essential for successful advanced information technology implementation. This is because advanced integrative information technologies like ERP and CRM are in essence process management tools and are evolving to become knowledge management tools.

Keywords: Enterprise Resource Planning, Customer Relationship Management, Quality Management, Organizational Learning

1. INTRODUCTION

Enterprise integration underlies e-commerce, enterprise resource planning, customer relationship management and other advanced uses of information technology to create business value [1-3]. Despite the proliferation of literature on e-commerce, the social and organizational aspects of enterprise integration are not well understood. Also, holistic approaches in enterprise integration research are rare. This paper aims to contribute to our existing understanding of e-commerce by proposing a framework that links organizational learning, quality management practices and the use of advanced information technologies such as enterprise resource planning and customer relationship management. The use of the framework will hopefully facilitate actions that create business value through the implementation of advanced information technologies. The theoretical framework has been derived from existing frameworks of organizational learning and quality management [4-12] and information systems implementation [13-15]. An assumption here is that information technology adoption would lead to better organizational effectiveness if it facilitates 1) practices that improved the quality of the product and services provided by the firm and 2) organizational learning.

Both quality management and organizational learning have been touted as approaches to improve organizational effectiveness [16-18]. While the link between quality management and organizational learning has been suggested by various authors [4, 9, 19], the nexus between information technology implementation, quality management and organizational learning is less pronounced.

2. THEORETICAL FRAMEWORK DEVELOPMENT

The framework has undergone three stages of development in the last eight years. First, it was developed to understand the link between organizational learning and quality management. The first stage of development involved a post-positivist qualitative and quantitative multi-level in-depth case study [20] of two organizations (herein called Netweb and Stark) that were fully owned subsidiaries of two multi-billion dollar (US\$) conglomerates. The first stage was conducted between early 1995 and early 1997. Then, the framework was adapted to understand the impact of information technology on the linkage between quality management and organizational learning in theoretical research in late 1998. Currently it has been recently adapted to understand the impact of advanced information technologies specifically enterprise resource planning and customer call center management (a subset of customer relationship management systems). The last stage involved a study of Stark within the wider context of the conglomerate to which it belonged using ethnographic approaches [21] in a multi-level multi-functional manner, and revisiting the evidence collected in phase I.

Netweb was a firm in the service industry and deployed advanced information and communicational technologies. The staff education levels of Netweb were comparatively high in relation to Stark. On the other hand, Stark was a manufacturing firm in a comparatively low technology industry. While Netweb experience high turnover rates being in a high tech industry, employees of Stark worked with the same company for relatively long periods, some even for decades. Netweb and Stark are summarized next.

Name	Netweb	Stark
Service/ Manufacturing	Service	Manufacturing
Technology	High	Low
Staff Turnover	High	Low
Staff Education	High	Low

Table 1 - Stark differences between two firms studied

The following table summarizes the fieldwork conduct of Phase I and Phase III. Note that Phase II involved only a theoretical adaptation of the framework developed in Phase I to account for the impacts of information technology implementation.

Phase I (positivist case study)	<i>No of org. levels studied</i>	<i>Number of Interviewees</i>	<i>Surveys</i>	<i>Other sources</i>
Stark (focused on MaxCo ET site of Stark)	5 levels (From Stark CEO to Machine Operator)	7 (Note: repeat interviews were conducted for some)	Organizational Learning Survey Learning Org. Profile	Observations, Meetings, documentation, informal conversations
Netweb (focused on the National Call Center)	5 levels (From Netweb CEO to call center operator)	10 (Note: same as above)	Same as above	As above
Phase II				
<i>Theoretical development. No field work in this phase</i>				
Phase III (ethnography)	<i>No of org. levels studied</i>	<i>Number of Interviewees</i>	<i>Surveys</i>	<i>Other sources</i>
Stark within context of the wider Solteria conglomerate	9 levels From former CEO of Solteria to Stark Truck Driver.	69 (Note: repeat interviews with many respondents)	Not applicable.	As above

Table 2 - Summary of conduct of fieldwork for Phase I, II and Phase III

From the table above, while efforts in Phase I involved testing a framework linking quality management and organizational learning, the critical ethnography in Phase III helped the author understand in a deeper manner the social and organizational aspects of enterprise systems implementation. This offered a very rich background against which to develop, question and extend theory [22]. For example, in phase III, over 40,000 pages of documentation were collected at Stark. Access was given to privy viewpoints that could not easily be gained using any other method except very in-depth case study methods (or ethnography). It should

be noted that Phase III started off as an interpretive ethnography to understand information technology-enabled enterprise integration. Constructs being studied were at that time ill-defined in the literature. Hence, ethnography was necessary.

The organizational learning survey (OLS) [10-12] and Learning Organization Profile (LOP) [7, 8] mentioned in previous table were instruments designed by other authors. They were among the few available instruments at the time the research was conducted (mid-1996 to early 1997) that assessed how the characteristics of an organization was in line with a learning organization [e.g., 17, 23]. It was not applied in phase III because the purpose of phase III was to achieve a deep understanding of the impact of enterprise resource planning on organizations. Moreover, Phase III began with the use of interpretive methods. Thus the use of survey methods was ruled out in phase III. Also, using a positivist framework at the start of the phase III was deemed inappropriate.

Combining the previous studies under a interpretive hermeneutic lens [24], allows one to conduct meta-studies that enrich current existing frameworks. Here, positivist frameworks are interpreted through an interpretive lens within the current context and applied to derive rich insights. This is because every act of research involves a double hermeneutic [cf. 25]. (Even positivist frameworks have to reconstructed and reinterpreted within new context every time they are applied [cf. 26]). From this process, an updated framework is derived that involves data of previous case studies and ethnographies being reinterpreted within a new framework presented in the next diagram. It is posited that advanced information technologies are involved in a structuration process of organizational practices, learning systems and the wider organizational environment for organizational learning. Also, quality management initiatives that are facilitated by information technologies (especially those that enable better process management capabilities such as customer relationship management and enterprise resource planning) impact on organizational practices, the learning systems and organizational environment for learning. The extent that information technology facilitates the linkages between actions that positively impact on the creation, maintenance and enhancement of organizational practices, learning systems and organizational environment that are conducive to organizational learning, is the extent that IT implementations may positively impact on organizational effectiveness and success. (The concept of success is dependent on who, when and how it is measured [27]. The author considers success to be related to the ability of the firm to implement information technologies to facilitate organizational learning. Also the framework has some overlaps between the areas of organizational practices, learning systems and organizational environment because they

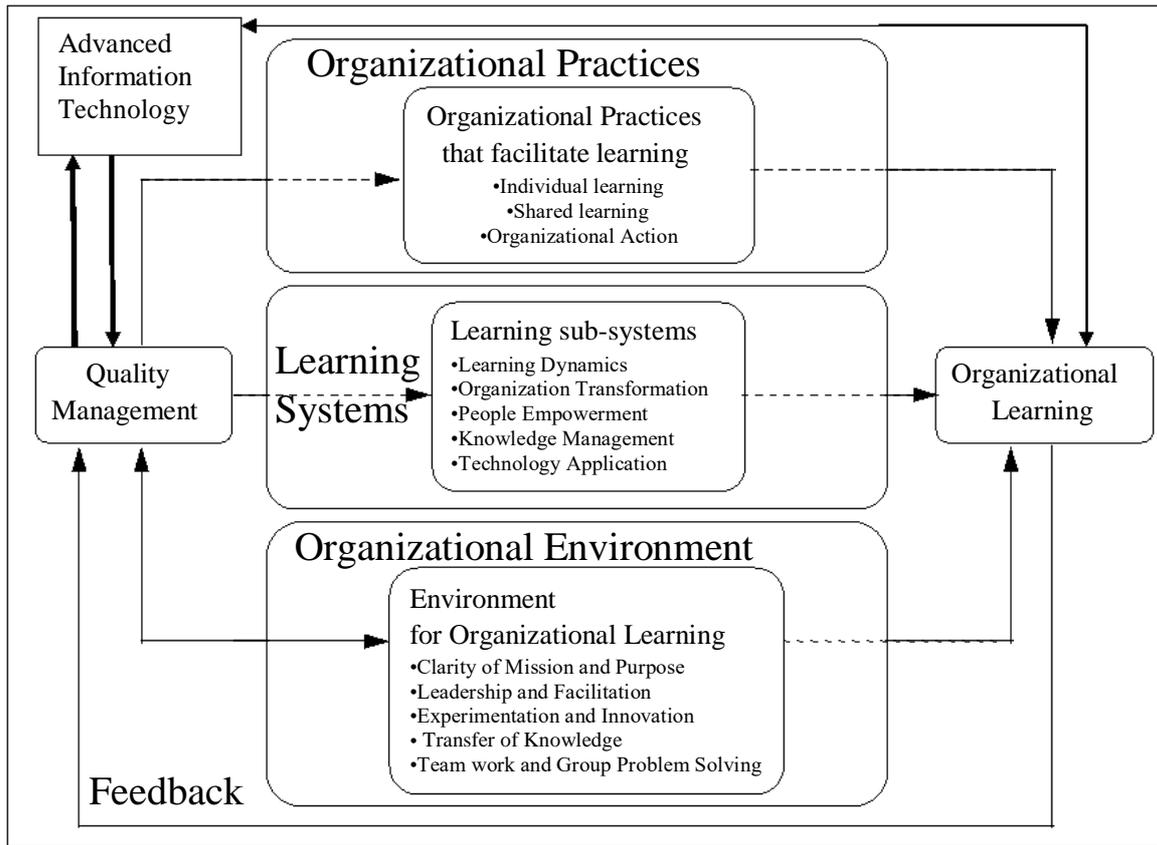


Figure 1 - Framework linking information technology, quality management practices and organizational learning

are not meant to be strictly independent or dependent variables but rather perspectives through which one understands organizational learning.

Organizational practices in the diagram above refer to institutionalized social practices that facilitate the linkages between individual learning, shared learning and organizational action [9]. Learning sub-systems refer to a perspective where several organizational sub-systems constitute a learning organization, namely learning dynamics, organizational transformation, people empowerment, knowledge management and technological application [7, 8]. Organizational environment for organizational learning is comprised of five characteristics: clarity of mission and purpose, leadership and facilitation, experimentation and innovation, transfer of knowledge and, team work and group problem solving [10-12]. Information technology through a structuration process may enable and constrain quality management initiatives that would then impact on organizational practices, learning systems and organizational environment. This then impacts on organizational learning that may also in turn impact the implementation of use of advanced information technologies. The use of advanced information technologies would also impact organizational learning directly through other means.

3. ANALYSIS

With Netweb, its call center was recognized widely as one of the best in the country at the time of research. When Netweb’s parent company started to expand its activities to another country, it showcased Netweb’s call center that had won a quality award that year as the epitome of excellence of its customer service. An examination of the social practices, learning systems and environment for organizational learning reveals that Netweb’s context was extremely conducive for organizational learning. Information technology used in this context facilitated the linkages referred above.

In contrast to this, Stark’s ERP implementation was not as successful. Although an environment conducive for organizational learning was created at Stark’s MaxCo ET site earlier, the overall organization and in particular, its order acceptance and delivery process was far from exhibiting characteristics of a learning organization. The ERP implementation was widely touted as a failure even after two years going live. This could be explained by the lack of the institutionalization of social practices that translated individual learning, to shared learning and eventually effective organizational action. The environment for organizational learning and the learning systems were also not conducive for organizational learning.

In the case of Netweb, the call center that was show-cased had strong Clarity of Mission and Purpose. The mission of the center was clearly defined and employees were clear about the purpose of the organization. There was clear Leadership modeled by the manager of the center who held monthly call center meetings where problems were resolved or escalated to more senior management if they had not already been resolved at the team level. However, with the Experimentation and Innovation, due to the center's emphasis on stability and predictability, there was not much emphasis put on devising new ways doing things. The Transfer of Knowledge was aided by the use of call center management software that recorded each problem that was logged with the center. This enabled all members of the center to view the problems that customers were experiencing and ensure all problems were resolved within the agreed time frames outlined in service level agreements. The environment for team work and group problem solving was facilitated by a team-based structure and rewards system which rewarded the teams as whole rather than just individuals.

The adoption of quality management principles by Netweb as whole institutionalized some of the aspects mentioned above. For example, the principle of Leadership modeling as prescribed by the Malcolm Baldrige criteria was adopted by the center's manager who was also an ardent supporter for Netweb's quality framework that was based on the Baldrige criteria.

With regards to learning sub-systems, Netweb had systems to facilitate Learning Dynamics. Single loop learning was facilitated by the adoption of a standard methodology for problem solving. However, at the senior executive level, organizational defensive routines hindered double loop learning [28] in the organization. With Organizational Transformation, the importance of being a quality oriented organization and a learning organization was understood and strongly supported by Netweb's senior management team. People Empowerment for learning was encouraged where managers and non-managers worked together to solve problems together. On the transfer of knowledge, Netweb's call center staff would store solutions to customer problems as quality system procedures so that important knowledge is coded, stored and made available to those who need and use it. Finally, Netweb among other things used Lotus Notes to manage group processes such as project management, team process and meeting management (an example of Technology Application to support organizational learning).

Organizational practices that facilitated learning were encouraged where for example; a customer logged a problem with the call center. This call is noticed by a customer service representative at the center. He solves that problem and writes a quality system procedure detailing the resolution procedure. So individual learning is then shared. The next time the same problem occurs, the quality system procedure to solve that

problem is applied. Thus there is a linkage between Individual Learning, Shared Learning and Organizational Action. Other practices that facilitated this linkage were weekly team meetings and monthly center meetings where problems and solutions were discussed and workers implemented these solutions.

Contrast this to Stark's call center. It was formed from an integration of the order acceptance and delivery processes of three of Stark's business units: MaxCo, CamCo and Xenon. For CamCo, the call center also merged the order acceptance and delivery processes that were previously localized at the level of each CamCo plant. This call center used an ERP system from a major vendor that was interfaced with a then state-of-the-art expert system for dynamic truck allocation. The dynamic truck allocation system recommended the best truck to delivery the order to the customer. However, due to the lack of Clarity of Mission and Purpose where a conflict of strategic orientations existed between the senior management, the call center did not fulfill its original purpose of ensuring on-time delivery of products. Certain dominant actors occasionally forced large orders from big customers through the system to maintain good relationships with these big customers; causing every order to be late. On Leadership and Facilitation, while the call center manager attempted to facilitate learning, the call center was part of the wider volatile centralized order acceptance and delivery cycle. Overall, there was no clear Leadership across this cycle that facilitated the environment for organizational learning. Experimentation and Innovation were constrained by frequent firefighting of problems. Furthermore, the transfer of knowledge across that order acceptance and delivery process was poor. For example, a sales person related to the author how the same problem of how add-ons to the product were not included for the third time in a row for the same customer! Furthermore, the plants that were integrated into the order acceptance and delivery cycle had previously competed against each other on the volume of product delivered. Thus teamwork and group problem solving were not practiced effectively.

With regards to learning sub-systems at Stark's call center, single loop learning was problematic where corrective actions to ensure the processes met their goals were not taken as discussed above in the case of missing add-ons to products. Double loop learning at the strategic level was also hindered. Refer to [29] for further information. Organizational transformation towards a learning organization was hindered by a change in leadership where the new CEO of Stark held different ideas of how best to generate value for the conglomerate that his predecessor. Furthermore, the senior manager who championed the learning organizational movement at Stark had already left. On people empowerment, Stark had a culture of cascading authority and autonomy to the lowest levels practicable. However, the call center impinged on prior established

autonomy and authority structures that involved delegating certain aspects of authority and autonomy to plant managers. Thus there was lack of cooperative behavior between managers throughout the order acceptance and delivery cycle. On knowledge management, although a Quality Improvement Reporting System was in place, it was not used effectively. There was a gap in the link between reporting a problem and ensuring it got fixed. On Technology Application for learning, Stark had implemented ERP, an expert system to assist in assigning the best truck to deliver products and a host of other then advanced information technologies. These supposedly could support a learning system. However, the use of these technologies to support a learning system was hindered due to the lack of overall process management throughout the order acceptance and delivery cycle that could have used data generated from these systems to improve performance.

From an organizational practices perspective, individual learning (such as the sales person discovering a problem described above) was not translated into shared learning (collective procedures and mindset to ensure that this problem does not occur) which is then translated into organizational action (the add-ons to the product for a customer are added and delivered correctly).

4. DISCUSSION

One could suggest that organizational learning is a criterion for effective advanced technology implementation such as enterprise resource planning and customer relationship management. For effective technology implementation, the applications would need to impact on processes to generate business value. However, as practice-based research shows, it is the tacit relationships between people and things that make processes work [30]. Moreover, these tacit relationships are highly contextual. The importance of improving social practices, learning systems and learning environment so that information technology supports linkages between quality management and organizational learning is underscored in this paper.

The areas outlined in the framework provide some key areas that practitioners and academics can focus on to ensure that implementation of advanced information technologies such as enterprise resource planning and customer relationship management contribute to a firm's performance. Theories, methods and tools established in other fields may be adopted to improve implementation success. For example, Argyris' ladder of inference and the left and right-hand column can be used to facilitate double loop learning [29].

5. CONCLUSION

It is posited that quality management and organizational learning principles and practices are essential for

successful advanced information technology implementation. This is because advanced information technologies like customer relationship management and enterprise resource planning are used to support process management [31] and will be increasingly used to support knowledge management. Thus principles, practices, frameworks, theories, methods and tools learnt in the last few decades from previous movements can surely be applied to ensure implementation success of these technologies. This paper has presented evidence to support the assertion that insights generated from the integration of quality management, organizational learning and information technology as illustrated in the framework presented in this paper could be used to explain if not improve IT-enabled enterprise integration success.

REFERENCES

- [1] L. M. Hitt, D. J. Wu, and X. Zhou, "Investment in Enterprise Resource Planning: Business Impact and Productivity Measures," *Journal of Management Information Systems*, vol. 19, pp. 71 - 98, 2002.
- [2] D. L. Goodhue, M. D. Wybo, and L. J. Kirsch, "The Impact of Data Integration on the Costs and Benefits of Information Systems," *MIS Quarterly*, vol. 16, pp. 293 - 311, 1992.
- [3] M. L. Markus and C. Tanis, "The Enterprise Systems Experience - From Adoption to Success," in *Framing the Domains of IT Research: Glimpsing the Future Through the Past*, R. W. Zmud, Ed.: Pinnaflex Educational Resources, Inc., 2000, pp. 173 - 207.
- [4] S. Shiba, A. Graham, and D. Walden, *A new American TQM: four practical revolutions in management*. Cambridge, Massachusetts: Productivity Press, 1993.
- [5] C. Argyris and D. A. Schon, *Organizational Learning: A Theory of Action Perspective*. Reading, Massachusetts: Addison-Wellesley, 1978.
- [6] P. Senge, "The Leader's New Work: Building the Learning Organization," *Sloan Management Review*, vol. 32, pp. 7 - 23, 1990.
- [7] M. J. Marquardt and A. Reynolds, *Global Learning Organization. Gaining Competitive Advantage through Continuous Learning*. Burr Ridge, Illinois: Irwin Professional Publishing, 1994.
- [8] M. J. Marquardt, *Building the Learning Organization. A Systems Approach to Quantum Improvement and Global Success*. New York: McGraw-Hill and American Society for Training

- and Development, 1996.
- [9] J. C. Lee, "The link between Total Quality Management and the Learning Organisation," presented at Australian and New Zealand Academy of Management Annual Conference, University of Wollongong, Australia., 1996.
- [10] S. C. Goh and W. Weisner, "Building a Learning Organisation: A Data Base and Focus Group Approach," presented at Human Resource Planning Society's 1995 Research Symposium, Ithaca, New York, 1995.
- [11] S. C. Goh and G. Richards, "Organizational Learning: A review and critique of the literature.," presented at Administrative Science Association of Canada Annual Conference, Halifax, N.S., 1994.
- [12] S. C. Goh and G. Richards, "The Organizational Learning Survey (OLS): Preliminary Scale Development and Validation Data," in *Unpublished Manuscript, Faculty of Administration, Univeristy of Ottawa*. Ontario, Canada, 1995.
- [13] W. J. Orlikowski and D. Robey, "Information Technology and the Structuring of Organizations," *Information Systems Research*, vol. 2, pp. 143 - 169, 1991.
- [14] W. J. Orlikowski, "Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations," *Organization Science*, vol. 11, pp. 404 - 428, 2000.
- [15] W. J. Orlikowski, "The Duality of Technology: Rethinking the Concept of Technology in Organizations," *Organization Science*, vol. 3, pp. 398 - 427, 1992.
- [16] G. Bounds, L. Yorks, A. M., and G. Ranney, *Beyond Total Quality Management: Toward the Emerging Paradigm*. Singapore: McGraw-Hill, 1994.
- [17] P. Senge, *The Fifth Discipline: The Art and Practice of Building the Learning Organisation*. Australia: Random House Publications, 1990.
- [18] R. Stata, "Organizational Learning-The Key to Management Innovation," *Sloan Management Review*, vol. 30, pp. 63 - 74, 1989.
- [19] P. Senge, "Building Learning Organisations," *Journal for Quality and Participation*, vol. 15, pp. 30 - 38, 1992.
- [20] R. K. Yin, *Case Study Research: Design and Methods*, 2nd ed. Newbury Park, California: Sage Publications, 1994.
- [21] M. D. Myers, "Critical Ethnography in Information Systems," in *Information Systems and Qualitative Research*, A. S. Lee, J. Liebenau, and J. I. DeGross, Eds. London: Chapman and Hall, 1997, pp. 276 - 300.
- [22] L. Harvey and M. Myers, "Scholarship and Practice: the Contribution of Ethnographic Research Methods to Bridge the Gap," *Information Technology & People*, vol. 8, pp. 13 - 27, 1995.
- [23] C. Argyris and D. A. Schon, *Organizational Learning II: Theory, Method and Practice*. Reading, Massachusetts: Addison-Wellesley, 1996.
- [24] H. Klein and M. Myers, "A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems," *MIS Quarterly*, vol. 23, pp. 67 - 93, 1999.
- [25] A. Giddens, *The Constitution of Society: Outline of the Theory of Structuration*. Cambridge: Polity Press, 1984.
- [26] R. E. Palmer, *Hermeneutics: Interpretation Theory in Schleiermacher, Dilthey, Heidegger and Gadamer*. Evanston, IL: Northwestern University Press, 1969.
- [27] M. Larsen and M. D. Myers, "When Success Turns into Failure: A Package-driven Business Process Re-engineering Project in the Financial Services Industry," *Journal of Strategic Information Systems*, vol. 8, pp. 395 - 417, 1999.
- [28] C. Argyris, *Overcoming Organizational Defenses. Facilitating Organizational Learning*. Needham Height, MA: Allyn & Bacon., 1990.
- [29] J. C. Lee and M. D. Myers, "Enterprise Systems Implementation Failure: The Role of Organizational Defensive Routines," Pacific Asia Conference on Information Systems, Shanghai, 2004.
- [30] J. S. Brown and P. Duguid, "The Social Life of Information." Boston: Harvard Business School Press, 2000.
- [31] T. H. Davenport, "The Future of Enterprise System-Enabled Organizations," *Information Systems Frontiers: special issue of on The Future of Enterprise Resource Planning Systems Frontiers*, vol. 2, pp. 163 -180, 2000.

Towards Integrated Document Management in Networked Product Development

Hannu Laesvuori, Paavo Kotinurmi

Software Business and Engineering Institute, Helsinki University of Technology, Finland
{hannu.laesvuori, paavo.kotinurmi}@hut.fi

Product development (PD) in a network of collaborative companies is different from the traditional way of doing PD within single organization. Parallel way of doing PD increases the need for synchronization of PD activities, but the flow of design information in the company network is difficult to manage. There are indications that the automation of document management should improve the performance of networked PD projects. However, the dynamic nature of networked PD is an obstacle for this, as it requires such flexibility and speed from the setup and further maintenance of integrated document management that is currently unavailable. This article is a work-in-progress publication about research that seeks ways to reduce the setup time of integrated document management for a network of companies that collaborate on PD. This article reviews literature to create the basis for developing solutions to solve this problem, presents our suggestions, and define the following research steps in the road towards integrated document management in networked PD.

Keywords: Product development, Product data management, PDM, integration, e-business, RosettaNet

1. INTRODUCTION

From business process perspective, e-business means automation of business processes by using electronic networks [25]. Vital for the automation of business processes is to ensure that the different information systems participating in the business processes can communicate with each other. E-business frameworks such as Electronic Data Interchange (EDI) and RosettaNet are collections of standards that can be used to achieve interoperability between the different information systems participating in a business process, thus enabling their communication [16][19][22].

The most important business processes of an enterprise are typically order fulfilment and product development (PD) [10]. Enterprises have been automating their order fulfilment processes for a considerable time, and e-business framework support for order fulfilment process automation exist. The benefits, such as reduced order fulfilment process costs, have also been reported [17]. PD process automation has progressed slower, and the e-business framework support for PD process automation is partly inadequate [12][13].

Previous research indicates that many PD projects have significant shortcomings, such as poorly functioning engineering change (EC) management that could be substantially improved by automated document management [2][10]. Besides the reported e-business framework support shortcomings for PD process automation, one important reason that prevents the automation of PD processes is that PD projects cannot afford as much time for setting up and maintaining the process automation, as order fulfilment processes. In networked PD the formation of the network may change on every new PD project, and often also during the PD project. This temporary nature of PD projects is a

fundamental difference in comparison to order fulfilment processes, which are typically continuous [10].

In this paper, we proceed to define the methodology for our work-in-progress research to support EC process automation in PD projects by enabling fast setup and configuration of Product Data Management (PDM) systems integration. The integration of PDM systems facilitates the automation of design document management, which has been found important in EC management and networked PD [2]. We also provide background information for the relevance of the research objective, and the results of a literature review that we conducted at the early part of the research.

Rest of this article is organized as follows. In section 2, we discuss certain networked PD project characteristics, and what role e-business frameworks play in automating PD processes. Section 3 presents our literature review. In section 4, we define our research protocol and a suggestion for a solution to enable faster setup of automated document management in networked PD. Section 5 presents our concluding remarks.

2. BACKGROUND

2.1. Product development in company networks

Recently companies have started to do PD in company networks instead of within a single organization. This is partly due to the increasing outsourcing of PD activities, but also because the shortening life cycles of products in many industry sectors, such as consumer electronics has increased the demand to reduce PD project lead-times [2]. The shift to do PD collaboratively in company networks instead of within a single organization has created problems. Organizational boundaries reduce the

informal communication between people involved in the PD project, and doing PD activities in parallel creates a need to synchronize the PD activities between all companies in the PD project. In addition, with multiple companies involved, the PD project is now being done in a more heterogeneous environment than before [2]. For example, when doing PD within a single company the design data can be managed with the company's PDM system. In networked PD, the functionalities offered by PDM systems, such as version control management, are typically not available as the design data is distributed across several different PDM systems that cannot interoperate [13].

Eloranta et al. [10] argue that project oriented business processes, such as PD, could benefit from partially automated document management conventions, and that the economic potential of the approach could be considerable. They argue that information sharing is the common denominator of all project oriented business processes and that the use of electronic documents would facilitate the information sharing.

The need to improve information sharing in networked PD has been reported in a case study of product geometry changes in injection moulding company network [2]. In this company network, there was a client company that developed consumer electronics products that required demanding plastic components developed by several supplier companies. The impact of product geometry changes to the initially estimated lead-time of the projects varied from 140% to 257%. According to the case study, PD processes are driven by the exchange of design documents such as CAD-models and improvement in the design document exchange process could have substantially reduced the impact of the product geometry changes. Two sources of extra work were found: working with an out-dated version of a design document or the absence of a design document. The case study suggested that the automation of design information flow between the enterprises could ensure better controllability and performance of the PD projects.

2.2. Support for automated document management in PD networks

Despite the advantages of automated document management in PD projects, it is not a common practice in the industry although a proof-of-concept demonstration exists [13]. PD in company networks involves collaborating with multiple companies that may change each time a new PD project is started. Thus, heterogeneity of the involved information systems, such as PDM systems, is inevitable and a solution supporting automated document management in PD networks should be based on standards.

Kotinurmi et al. [12] argue that e-business frameworks are the standards that could support automated

document management in PD networks. Approaches such as point-to-point integration and portal were considered infeasible, as portals require more human involvement and the management of point-to-point integration is not feasible in a large company network. Kotinurmi et al. evaluated 15 different e-business frameworks based on their support for process integration, secure messaging, PDM related messages and industry usage. Based on the evaluation, RosettaNet was considered best suited to support automated document management in PD projects.

The most important part of the RosettaNet standard are the inter-company business process definitions called Partner Interface Processes (PIP). RosettaNet marks each PIP with maturity level information, and the highest maturity level has been assigned to several order fulfilment PIPs. PIPs for PD processes typically have a lower maturity level, and although the PIP category "Collaborative Design & Engineering" existed already in the beginning of year 2002, it still contains not even drafts [20]. This PIP category would likely contain the PIPs for enabling automated document management that are currently missing from RosettaNet [12]. Thus, even if RosettaNet is the best-suited e-business framework to support networked PD, it lacks process definitions to support it [13].

Case studies of RosettaNet implementations, such as the Arrows case study conducted by the Stanford University, suggest implementation times measured in months [20]. Indeed, even with mature vertical industry standards, it is necessary to agree with trading partners (TP) on how to use the standard, which can take considerable time [6]. An implementation time measured in months is clearly infeasible for a PD project that may last only a few months. It appears that networked PD requires different approach to the process automation implementation than has been used before.

3. LITERATURE REVIEW

This section presents our literature review process and the methodological background. We classify the key articles found in the literature review, and present a short summary of each.

Literature review was performed to find existing scientific knowledge on e-business architectures, PD information systems integration and the requirements for IT support in networked PD projects. Research was carried out according to the structured approach recommended by Webster and Watson [28].

The first step of literature review was performed by looking through the relevant top journals of Computer Science and Management Information Systems. Our choices were MIS Quarterly, Communications of the ACM, Computers in Industry, Production Planning & Control, Journal of Manufacturing Technology

Management (formerly Integrated Manufacturing Systems), Information Systems Research, Journal of Systems and Software, and Computer Standards & Interfaces. Due to recent upraise in e-business and business process automation, we examined the journals starting from year 2001. The full text of each article was reviewed in order to eliminate those articles that were not related to our research problem.

From the key articles found from the journals, we reviewed the citations for the articles. The key articles were also forward examined to find more articles for the review. Using these methods, we found a few more articles. In addition, we defined a set of keywords such as 'e-business architecture' and 'PDM integration'. Then we searched a number of electronic collections such as Science Direct, the ACM digital library, IEE Explore, Citeseer, Web of Science and Lecture Notes in Computer Science using the keywords for full text searches. This led to a discovery of few new articles relevant to our research. Validation was accomplished by comparing the articles we found against those we had found previously by unsystematically observing research publications on the field. The literature review had found all important articles and many new ones, 48 articles in total.

3.1 Fast collaborative network setup, research on virtual enterprises

The research on virtual enterprises (VE) and virtual organisations has covered issues related to integrating participating enterprise information systems within virtual enterprise partners. Virtual enterprise, defined as a network of autonomous firms that co-operate to achieve common business goals, has received considerable attention in the academic community [4][14]. Although multiple software architectures to support virtual enterprises have been defined, e.g. ones listed in [4][14][30], there are still many challenging issues requiring further research [4]. These include collaborative network support for setting up networks, operating them, and the evolution and dissolution of the network. In addition, definition of ICT infrastructures and identifying relevant standards that enable co-operation are needed. E-business frameworks are expected to have an impact on the development of support software for virtual organisations and other collaborative networks [5]. The use of Standard for the Exchange of Product model data (STEP) and Internet communication have also been proposed to support the data exchange between the different systems used at the different companies participating in the VE [30].

3.2 E-business architectures

Several publications present software architectures supporting e-business [1][3][4][7][14][16][23][24] utilizing XML. However, most of them are only suggestions without reporting experiences on

implementation and use. Only a few experience papers using modern XML-based e-business frameworks exist [18][19][21].

3.3 PD collaboration and PDM integration

Liu and Xu [15] state requirements for web-based PDM systems supporting collaboration. They propose web browser PDM interface as a solution. The modern PDM systems have web interfaces, but they differ in usage logic. This is why integration is needed also between participating systems. For PDM integration, agent-based systems have also been developed [27]

There are many collaboration infrastructures for PD collaboration based on the use of Common Object Request Broker Architecture (CORBA) and STEP [8][9][29][30]. The modelling capabilities distributed over the Internet using CORBA facilitate collaboration between product designers, and STEP provides the common terminology. These papers lack industrial validation, and most concentrate just on laboratory testing. The use of STEP in industry is limited, concentrating in geometry based applications.

4. RESEARCH PROTOCOL

In a research project conducted at Helsinki University of Technology, we are trying to improve the performance of engineering change (EC) processes in networked PD. A key element in this is to use standard, automated inter-company processes between the members of the PD network. We have identified two obstacles for this: 1) standard process definitions for inter-company engineering change management do not exist 2) current approaches to achieve systems integration are too slow. The two obstacles are inter-connected, as the process definitions and their use set requirements for the systems integration needed to enable them. We are working on the process definitions and currently they are based on a case study of two separate PD networks.

In this paper, we define the methodology for our work-in-progress research to address the second obstacle, i.e. how to setup integrated document management for new, networked PD projects, in a short time.

The rest of this section is organized as follows. First we give a brief overview of the design research methodology on which our research protocol is based. Then, we proceed to identify how the different design research phases relate to our research.

4.1. Design research

Our research approach is based on the design research methodology [11]. Design research views the creation of new knowledge as a loop, as illustrated in figure 1. The loop starts from awareness of a problem, followed by a suggestion how to solve it. The suggestion is then

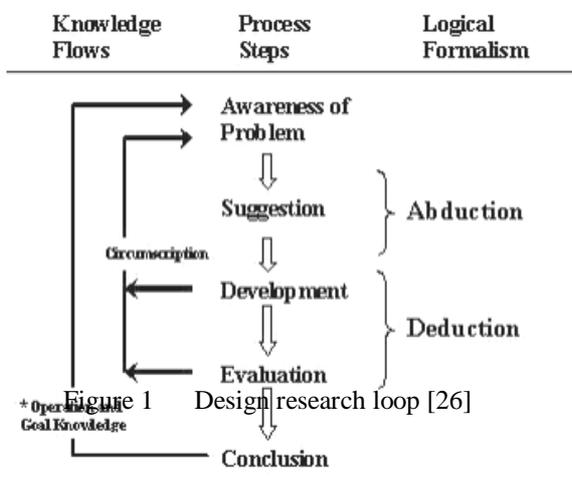
developed, and the result is a new artifact. The artifact is then applied to the problem, and the results are evaluated. If the artifact solved the problem, the conclusions present new knowledge that can be used on subsequent research. If it did not solve the problem, a new suggestion, development and evaluation follow.

In design research, the suggestion is usually done intuitively, but it must always be novel and innovative, either solving a new problem or solving an old problem in a more effective way than previously. Moreover, the created artifact must always be clearly defined, evaluated and reported [11] [26].

The position of our research in the design research loop is currently in the initial stage of development. We are now aware of the problem and have a high-level suggestion on how it might be solved. We expect that the development work will proceed iteratively together with a more detailed solution suggestion.

4.2. Awareness of the problem

In our case, the design research loop started with the awareness of the research problem through talks with several industry experts in previous PD related research projects. Next, we reviewed existing knowledge by conducting a literature review to confirm that we are facing a new and valid research objective. The results of the literature review, and the commitment from the companies represented by the industry experts to



participate in the research, convinced us that we are facing a real research problem.

4.3. Suggestion

The project-oriented nature of networked PD would require repeatable systems integration work, if done in the traditional way. Because of the inevitable ambiguity of standard specifications and how different individuals understand them, and the diverse needs of companies using the standards, PDM systems integration using e-business frameworks will likely continue to consume considerable time in the near future [6][13][20].

In our opinion a solution to support fast setup of systems integration might be based on the concept of reusing past systems integration work between different PD projects. This seems feasible, as networked PD is usually done within fairly small and stable network of companies, with only the roles and project teams changing frequently. Thus, only the first systems integration work between two TPs collaborating in networked PD would require considerable time. Systems integration work for subsequent networked PD projects between the two TPs would be small.

The limited size of the network and relative stability of the partnerships in networked PD, is different from the requirements for VEs. However, there are still many commonalities between the requirements for networked PD and VE, so the research for VE might be used in our research.

4.4. Development

We are now in the early phase of the development of an artifact to enable fast integration of PDM systems in networked PD. The artifact will consist of a concept, a software tool, supporting software architecture and a management process.

The reuse of past systems integration work between two TPs in future networked PD projects would in our opinion require dividing the parameters of PDM systems integration in two groups: PD-project related parameters and TP related parameters. For example, parameters such as 'Document Type' that derive from the PDM systems that the TPs use, or IP-addresses required for data exchange are likely to remain same in all networked PD projects between the two TPs. However, parameters, such as project identifiers and contact persons for exceptional situations, typically change on every project.

These two groups of parameters must be defined, as not only is their speed of change different but typically they are controlled by different parties. For example, the TP related parameters may be controlled by people with technical background who are responsible for the IT infrastructure of the company. The project related parameters, however, may be controlled by a project manager in a PD group. People with different backgrounds may require different software tools, or at least user-interfaces, for managing the parameters. Moreover, if the people reside in different divisions of the company, it may be obstacle for the information flow between them.

Previously these two groups of parameters have been treated as one. B2B integration architectures such as the one presented by [3] recognize the need for TP management, but not for project data management. In our experience, the software tools for configuring systems integration parameters are suited for

configuring TP parameters, but not for project related parameters.

4.5 Evaluation and conclusions

The artifact will be iteratively designed and evaluated by an industrial focus group. Once this preliminary validation has been achieved, the adoption of the artifact by industry would bring stronger market validation. The artifact and results of the evaluation will be reported publicly, and evaluated against the other related research we found during the literature review.

5. CONCLUDING REMARKS

Previous research indicates that engineering change (EC) management in a network of companies collaborating on product development (PD) is often functioning poorly. Automation of design document exchange process between the companies could offer considerable potential for savings, but it has been found difficult. The use of e-business frameworks, particularly RosettaNet, in the automation of the design document exchange process has been suggested.

In this paper, we argue that there are two obstacles for using RosettaNet in the integration of the design document exchange process: unsuitable or missing process definitions for EC management, and typical implementation times that are too long for networked PD.

We define the methodology for work-in-progress research to seek ways to reduce the setup time of automated document management for a network of companies collaborating on PD. Our approach is to introduce basis for reuse in systems integration work, so that subsequent PD projects between two trading partners could benefit from past integration work.

The contributions of this article are the definition of the research objectives, the methodology for solving it, a potential solution suggestion and the literature review.

The literature review indicated that project oriented business processes, such as PD, would benefit from increased integration, but that e-business frameworks have not been previously specifically discussed within PD needs as older standards such as STEP and CORBA have dominated the research applications. We have found previous research that can be compared to our approach, and can aid to position our research in relation to other research on the subject.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support from the National Technology Agency of Finland (Tekes) and the support of Graduate School for Electronic Business and Software Industry (GEBSI).

REFERENCES

- [1] Ball, M., Ma, M., Raschid, L., Zhao, Z., "Supply Chain Infrastructures: System Integration and Information Sharing", *ACM SIGMOD special issue: Data management issues in electronic commerce*, Vol. 31, No. 1, pp. 61–66, 2002.
- [2] Borgman, J., Sulonen, R., "A Case Study of the Impacts of Preliminary Design Data Exchange on Networked Product Development Project Controllability", *Proc. of the Int. Conf. on Engineering Design*, Stockholm, 2003.
- [3] Bussler, C., "The role of B2B engines in B2B integration architectures", *ACM SIGMOD special issue: Data management issues in electronic commerce*, Vol. 31, No. 1, pp. 67–72, 2002.
- [4] Camarinha-Matos, L., Afsarmanesh, H., "Elements of a base VE infrastructure", *Computers in Industry*, Vol. 51, No. 2, pp. 139-163, 2003
- [5] Camarinha-Matos, L., Tschammer, V., Afsarmanesh, H., "On Emerging Technologies for VO" in Camarinha-Matos, L. M.; Afsarmanesh, Hamideh (Eds.) *Collaborative Networked Organizations A research agenda for emerging business models*, pp. 207-224, 2004.
- [6] Casati, F., "A Conversation on Web Services: what's new, what's true, what's hot. And what's not.", *Proceedings of the ECAI workshop on Knowledge Transformation and the Semantic Web*, Lyon, 2002.
- [7] Chan, S., Dillon, T., Siu, A., "Applying a Mediator Architecture Employing XML to Retailing Inventory Control", *The Journal of Systems and Software*, Vol. 60, No. 3, pp. 239–248, 2002.
- [8] Chung, J., Lee, K., "A framework of collaborative design environment for injection molding", *Computers in Industry*, Vol. 47, No. 3, pp. 319–337, 2002.
- [9] Domazet, D.R., Miao Chun Yan, Chee Fook Yew, Kong, H.P.H., Goh, A., "An Infrastructure for Inter-Organizational Collaborative Product Development", *Proc. of the 33rd Hawaii Int. Conf. on System Sciences*, pp. 2176–2185, 2000.
- [10] Eloranta, E., Hameri, A-P., Lahti, M., "Improved project management through improved document management", *Computers in Industry*, Vol. 45, No. 3, pp. 231-243, 2001.
- [11] Hevner, A., March, S., Park, J., Ram, S., "Design Science in Information Systems Research", *MIS Quarterly*, Vol. 28, No. 1, pp. 75-105, 2004.
- [12] Kotinurmi, P., Borgman, J., Soininen, T., "Design Document Management in Networked Product Development Using Standard Frameworks", *Proc. of the Int. Conf. on Engineering Design*, Stockholm, 2003.
- [13] Kotinurmi, P., Laesvuori, H., Jokinen, K., Soininen, T., "Integrating Design Document Management Systems using the RosettaNet E-business Framework", *Proc. of the Int. Conf. on Enterprise Information Systems*, pp. 502-509, Porto, 2004.

- [14] Kovács, G., Paganelli, P., "A planning and management infrastructure for large, complex, distributed projects—beyond ERP and SCM", *Computers in Industry*, Vol. 51, No.2, pp. 165-183, 2003.
- [15] Liu, D.T., Xu W. P., "A review of web-based product data management systems", in *Computers in Industry*, Vol. 44, No. 3, pp. 251-262, 2001.
- [16] Medjahed, B. Benatallah, B., Bouguettaya, A., Ngu, A. H.H., Elmagramid, A.K., "Business-to-business interactions: issues and enabling technologies", *VLDB Journal*, No. 12, pp. 59-85, 2003.
- [17] Mukhopadhyay, T., Kekre, S., Kalathur, S. "Business Value of Information Technology: A Study of Electronic Data Interchange", *MIS Quarterly*, Vol. 19, No. 2, pp. 137-156, 1995.
- [18] Nurmilaakso, J.M., "XML-based Supply Chain Integration: A Review and a Case Study", *Licentiate thesis in Helsinki University of Technology*, 2003.
- [19] Nurmilaakso, J.M., Kotinurmi, P., "A review of XML-based supply-chain integration", *Production Planning & Control*, Vol. 15, No. 6, pp. 608 – 621, 2004.
- [20] RosettaNet, "RosettaNet standards", <http://www.rosettanet.org/standards>, August 2004.
- [21] Sayal, M., Casati, F., Dayal, U., Shan, M-C., "Integrating Workflow Management Systems with Business-to-Business Interaction Standards", *Proc. of 18th Int. Conf. on Data Engineering*, pp. 287 –296, 2002.
- [22] Shim S.S.X., Pendyala, V.S., Sundaram, M., Gao, J.Z., "Business-to-Business E-Commerce Frameworks", *IEEE Computer*, Vol. 33, No. 10, pp. 40-47, 2000.
- [23] Shin, K., Leem, C.S., "A reference system for internet based inter-enterprise electronic commerce", *Journal of Systems and Software*, Vol. 60, No. 3, pp. 195-209, 2002.
- [24] Sundaram, M., Shim, S., "Infrastructure for B2B Exchanges with RosettaNet", *Proc. of Third Int. Workshop on Advanced Issues of E-Commerce and Web-Based Information Systems*, pp. 110–119, California, 2001.
- [25] Turban, E., King, D., Lee, J., Viehland, D., "Electronic Commerce, a Managerial Perspective", *Pearson Education, Inc.*, 2004.
- [26] Vaishnavi, V., Kuechler, W., "Design Research in Information Systems", <http://www.isworld.org/Researchdesign/drisISworld.htm>, February 2004.
- [27] Wang, Y.D., Shen, W., Ghenniwa, H., "WebBlow: a Web/agent-based multidisciplinary design optimization environment". *Computers in Industry*, Vol. 52, No. 1, pp. 17-28, 2003.
- [28] Webster, J., Watson, R. "Analyzing the Past to Prepare for the Future: Writing a Literature Review", *MIS Quarterly*, Vol. 26, No. 2, pp. xiii-xxiii, 2002.
- [29] Zhang, Y.D., Shen, W., Ghenniwa, H., "A review of internet-based product information sharing and visualization". *Computers in Industry*, Vol. 54, No. 1, pp. 1-15, 2004.
- [30] Zhang, Y., Zhang, C., Wang, P., "An Internet based STEP data exchange framework for virtual enterprises", *Computers in Industry*, Vol. 41, No. 1, pp. 51-63, 2000.

Transaction Governance Structures of Brick-and-Mortar and Click-and-Order: An Instrument Design Study

Aimao Zhang

Department of Information Technology, Georgia Southern University, P. O. Box 8150, Statesboro, GA
30460-8150, USA {aimao}@georgiasouthern.edu

ABSTRACT

Since the inception of the construct of transaction governance structure (TGS), transaction cost economics (TCE) has become an important anchor for the analysis of a wide range of economic and organizational issues. As theory of TCE advances, the perception of TGS has shifted from a polar classification (market/hierarchy) towards a continuum of market-hierarchy [17][20]. Despite the development in conceptual framework, empirical work based on the idea of the market-hierarchy continuum is scarce. Part of the difficulty is the lack of clear defined and operational dimensions of TGS. Although dimensionalization of transaction has received early and explicitly attention, the dimensionalization of TGS is relatively limited. This paper is an initial effort in instrument building. In this paper, we will (1) review the literature on TGS, (2) survey 40 empirical studies from 1982 to 2002, (3) present dimensions of TGS based on Williamson's work of 1991, (4) run a preliminary test to compare TGSs of brick-and-mortar and click-and-order.

Keywords: Transaction governance structure, industrial economics, contract law, e-commerce.

1. INTRODUCTION

In the recent years, we have observed the broad shifts of TGS in the United States such as a large-scale integration among banks, security firms, and insurance companies [34] and a continuous migration toward market TGS in the computer industry in which component manufacturers are replacing fully-integrated computer manufacturers [23]. This confusing landscape of TGS changes raises questions about reasons for the TGS shift, the nature of the new TGSs, and the influential factors in designing and selecting a TGS. At the macro level, the answer to these questions will help policy makers in developing economic and governmental policies. At the micro level, the answer to these questions will help company managers to make decisions regarding TGS, such as what to buy, where to buy, or how to buy.

The fundamental construct for answering these questions is TGS. Since the inception of the construct of transaction governance structure (TGS), transaction cost economics (TCE) has become an important anchor for the analysis of a wide range of economic and organizational issues. As theory of TCE advances, the perception of TGS has shifted from a polar classification (market/hierarchy) towards a continuum of market-hierarchy [17][20]. Despite the development in conceptual framework, empirical work based on the idea of the market-hierarchy continuum is scarce. Part of the difficulty is the lack of clear defined and operational dimensions of TGS. Although dimensionalization of transaction has received early and explicitly attention, the dimensionalization of TGS is relatively limited. This paper is an initial effort in instrument building. The organization of the paper is as follows. Section 1 is an introduction. Section 2 reviews the literature on TGS. Section 3 surveys 40 empirical

studies from 1982 to 2002. Section 4 presents dimensions of TGS based on Williamson's work of 1991. Section 5 runs a preliminary test to compare TGSs of brick-and-mortar and click-and-order. Section 6 is a conclusion.

2. LITERATURE REVIEW

2.1 Transaction Cost Economics (TCE)

The construct of TGS is defined as the structure that mediates exchanges of goods or services between technology separable entities—businesses, subdivisions of a business, or individual buyers and sellers [46][47]. Ring and Van de Ven further defined TGS as “the legal forms of governance that apply to different kinds of transactions (ranging from markets to hierarchies), and the structural and procedural safeguards that parties negotiate into a transaction” [38]. Traditional TCE perceives TGS from a dichotomous view in which market and hierarchy are mutually exclusive [11][47]. At one end lie purely market-based TGSs, in which price is the invisible hand that mediates supply and demand [40]. At the other end, one finds purely hierarchical TGS, in which a central authority controls and allocates resources [9]. Typically, attempts to explain the selection of a TGS for a particular good or service have been based on comparisons between market and hierarchy transaction costs [11][46][47]. Generally speaking, when asset specificity of a transaction is low, market TGS is more economical and preferable. When asset specificity is high, hierarchy has the advantage of lower governance costs. A substantial amount of research has produced results consistent with transaction cost theory [4][26][27][42][44].

2.2 Network

TEC rests on the premise that market and hierarchy are mutually exclusive. Powell [35] does not agree with this premise or the concept of a market-hierarchy continuum, which is suggested by John and Reve [20]. Powell describes the network as the third independent form of governance structure. Ouchi [30] claims that price, rule, and trust are control mechanisms for markets, hierarchies, and networks, respectively. Bradach and Eccles [7] argue that three control mechanisms (price, authority, and trust) can be combined in a variety of ways under each structure, e.g., price and authority mechanisms under a market structure or a price and authority combination in a hierarchy.

3. REVIEW OF 40 EMPIRICAL STUDIES 1982-2002

This paper reported 40 empirical studies on TGS from 1982 to 2002. These studies had examined the construct of TGS either as an independent variable or a dependent variable. The review of these studies was focused on the dimensionalization of TGS and the type of instrument developed (see Table 1, available from the author upon request).

The majority of studies in 1980's perceived TGS as dichotomous and developed single-dimension and single-item scales [1][2][3][4][43][44][27]. These dichotomous single-dimension and single-item scales used 1 and 0 to record the decisions such as buy or make, integrate or not integrate, use internal work force or use external agents. Following that, some multichotomous scales were developed [5][15][21][22]. These multichotomous single-dimension and single-item scales measured TGSs into discrete categories (market, joint venture, partner relationships, hierarchy) along the market-hierarchy continuum. In 1990's, as the perception of TGS shifted from a polar classification (market/hierarchy) towards a continuum of market-hierarchy, multi-dimensional and multi-items instruments were developed to incorporate the variety of cooperative relationships [8][17][19][20][18][29][32][33]. Table 1 categorized the 40 empirical studies into groups according to the types of scale used.

There were two interesting trends among the multi-dimensional instruments. First, several multi-dimensional instruments captured the dimensions of formalization, centralization, flexibility and control [8][19][20]. Such dimensionalization gave a flavor of organization theory from Pugh who provided the theoretical framework for study organization structure [36][37]. According to Pugh and others, "six primary dimensions of organization structure were defined: (1) specialization, (2) standardization, (3) formalization, (4) centralization, (5) configuration, (6) flexibility" [37].

Secondly, dimensions of information exchange and interaction started to appear in middle 80's [18][29][31][32][33]. This trend was coincided with the time period when global competition intensified, firms moved towards flatter and more horizontal structures, and management emphasized cooperation rather than competition among self-managed teams and among inter-organization alliances [12]. In the same time period, we had seen the new research streams in managing dynamic processes [14], social network analysis [35], and information processing and conflict management [10][28].

4. DIMENSIONS OF TGS

4.1 Williamson's Framework

There was no consistence on the dimensions of TGS among the previous studies. Different scholars emphasized different dimensions of the construct. In defining dimensions of TGS, this study is going to follow the dimensionalization framework put forward by Williamson [45]. Williamson's framework was a recent contribution to TCE studies, and it incorporated dimensions of TGSs arrayed on the entire range of the market-hierarchy continuum. Williamson identified the key dimensions that differentiate three generic forms of TGS (market, hybrid, hierarchy). These dimensions are incentive, control mechanisms, adaptabilities, and types of contract law. The incentive is the degree of the price acting as an incentive to trigger and control actions. Administrative controls are the level of control and manipulative power gained through monitoring, career rewards, and penalties. Adaptation (A) is the ability to take the right actions to adapt to the new equilibrium whenever there are changes in the demand or in supply. Producers and buyers independently and autonomously response to the price changes to maximize their profits. Adaptation (C) is the ability to coordinate among interdependent parties to align with a collective goal. According to Williamson [45], firms, hybrid, and hierarchy are governed under different laws—classical contract law, neoclassical contract law, and forbearance respectively. The role of court becomes less important from classic to neoclassic to forbearance law. Table 2 (available from the author upon request) exhibits how market, hybrid, hierarchy are different along these dimensions.

4.2 Fit the 40 studies into William's Framework

In Table 3, the existing dimensions from 40 empirical studies were assigned into the five dimensions defined by Williamson. Although the assignments were subjective and crude, this exercise was helpful in mapping relationships and identifying the dimension that needed further development. As you can see from Table 3, the incentive dimension was underdeveloped. There was limited number of item designed for identifying suppliers' or buyers' price sensitivity or

readiness of adjusting quantity when price and demand fluctuated.

Table 3. Assignment Of The Existing Dimensions Into Williamson’s Framework

Incentive through price <ul style="list-style-type: none"> • Exclusivity [8] • Pattern of payoffs [32]
Administrative control <ul style="list-style-type: none"> • Monitoring [41] • Buyer’s control over supplier decision making [17] • Centralization [21] • Centralization [20] • Centralization [19] • Control [19] • Financial incentives [8] • Monitoring of supplier [29] • Operating controls [33]
Adaptation (A) <ul style="list-style-type: none"> • Replaceability of commission income [17] • Exit costs [40-A] • Exit barriers [8] • Restraint in the use of power [18]
Adaptation (C) <ul style="list-style-type: none"> • Collaboration [40A] • Collaborative interaction [40-A] • Joint action [17] • Interactions [20] • Behavioral transparency [32] • Frequency of interaction [32] • Flexibility [18] • Information exchange [18] • Shared problem solving [18] • Supplier flexibility [29] • Supplier assistances [29] • Information provided to supplier [29] • Exchange information [33] • Sharing benefit TGS and burdens [33] • Adaptation [31] • Types of adjustments [31] • Information exchange for long-term planning [31] • Information exchange for structural planning [31]
Contract law <ul style="list-style-type: none"> • Buyer’s and supplier’s commitment to the relationship [4A] • Formalization [21] • Expectations of continuity [17] • Formalization [20] • Formalization [19] • Formality [8] • Time horizon [32] • Expectation of continuity [29] • Contractual focus [33] • Relationship focus [33] • Method of enforcement [31]

5. A PRELIMINARY TEST

5.1 Instrument development and test

The emergence of e-commerce has provided a variety of choices of TGS such as online exchanges, distributor consortia, supply chain networks, auction, and mass catalog compilers [6]. TEC can help us answer the questions of what product is suitable for E-commerce (digital or tangible products), where to buy (brick-and-mortar or click-and-order), or how to buy (auction or private network). In this study, we will conduct a preliminary test on TGSs of brick-and-mortar and click-and-order.

A panel of expert has generated 24 items (see Table 4) along the five dimensions recommended by Williamson [45]. All items were presented with 5-point Likert scale ranging from “strongly disagree” to “strongly agree”. Each item is applied to “online” and “offline” separately and respondents checked their level agreement accordingly. One dependent variable was coded 0 for online and 1 for offline, and five independent variables were summated scales of five dimensions. Each summated scale equaled to the average of measurements of all items of the same dimension. The respondents were college students. Out of all respondents, 20% had shopped online at least 5 times, 30% had shopped online at least 10 times, and 50% has shopped online at least 20 or more times. The sample of 124 was divided into an analysis sample of 102 observations with the remaining 22 observations constituting the holdout or validation sample.

5.2 Result

We performed logistic regression on the data set based on the equation below [16].

$$p(y = 1 | x) = \frac{e^{ax+c}}{1 + e^{ax+c}} \tag{1}$$

Where:

$$ax = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5;$$

e is the base of natural logarithms;

y is dependent variable (0, 1);

*x*₁, *x*₂, *x*₃, *x*₄, *x*₅ is summated variables representing incentive, administrative control, adaptation(A), adaptation(C), contract law respectively.

*a*₁, *a*₂, *a*₃, *a*₄, *a*₅ were obtained from regressing result (Table 5, available from the author upon request). Thus *ax* = .187*x*₁ + 2.503*x*₂ + 1.799*x*₃ + .692*x*₄ + 1.349*x*₅ - 19.297.

Table 4. Items For Comparing TGSs Of Brick-And-Mortar And Click-And-Order

Incentive <ul style="list-style-type: none"> • There is a big price discrepancy among the same type of products offered by different sellers. • I do a lot of price comparison among different sellers when I’m shipping.*
--

Administrative control
<ul style="list-style-type: none"> • A seller has influence over a buyer's decision-making. • A buyer has influence over a seller's decision-making. • A seller does not know very much about you as a buyer.* • You as a buyer do not know very much about a seller.* • A seller can monitor your shopping behaviors. • A seller can manipulate your shopping behaviors. • It is easy for a seller to cheat and misrepresent a product.*
Adaptation (A)
<ul style="list-style-type: none"> • A seller adjusts his price and production quantity according to demand.* • Fluctuation of market demand has little impact on price. • There are many suppliers making the same or similar products.* • The goal of a seller is to maximize his profit regardless of loss of gain of any other players.* • The goal of a buyer is to minimize his cost regardless of loss of gain of any other players.* • It is easy to become a seller in terms of initial investment and market entry cost.*
Adaptation (C)
<ul style="list-style-type: none"> • Negotiations or discussion between a buyer and a seller is required to complete a transaction. • Explicit communication is required to coordinate activities among sellers and buyers. • A seller and a buyer usually settle a problem by cooperation and negotiation. • I would negotiate and cooperate with a seller.
Contract law
<ul style="list-style-type: none"> • A supplier and a buyer usually go to a court to settle a dispute.* • A supplier and a buyer usually settle a dispute through discussions, meetings, or arbitration. • A seller usually has to obey the rules and laws set up by a local community.
* Reserved.

Hosmer and Lemeshow test measured the correspondence of the actual and predicted values of the dependent variable. A better fit would be indicated by a smaller difference in the observed and predicted classification or by a non-significant chi-square value. Our chi-square value was 10.093, which is not significant because $p=0.259$. Table 6(available from the author upon request) presented the classification result. In the analysis sample, 74% percent of cases were predicted correctly. In the holdout sample, 63% of cases were predicted correctly.

6. CONCLUSION

In summary, perception of TGS advanced from dichotomous to continuous [17][20], from single dimension to multiple dimension [29], from local focus of by-or-make decision [43] to global discussion of international alliances [21][22]. We also observed the influence from other disciplines on TCE, such as organization theory [16A] and information theory [13].

“Big ideas often take a long time to take on definition. Thirty-five years passed between Coase's 1937 article

and efforts to operationalize transaction costs in the early 1970's (Williamson, 2000. pp.31)” by [46] and another twenty years passed before Williamson dimensionalized the construct of TGS [45]. The previous theoretical and empirical works have laid a solid foundation for us to bring the research forward. The future research is to conduct empirical studies to verify the construct of TGS, to apply TEC in analyzing the whole range of cooperative relationships, to move TEC beyond traditional commerce into e-commerce.

REFERENCES

- [1] Anderson, E., “Transaction costs as determinants of opportunism in integrated and independent sales forces”, *Journal of Economic Behavior and Organization*, Vol. 9, No. 3, pp247-264, 1988.
- [2] Anderson, E., “The salesperson as outside agent or employee: A transaction cost analysis”, *Marketing Science*, Vol. 4, No. 3, pp234-254, 1985.
- [3] Anderson, E., A. T. Coughlan, “International market entry and expansion via independent or integrated channels of distribution”, *Journal of Marketing*, Vol. 51, No. 1, Jan; pp71-82, 1987.
- [4] Anderson, E., D. C. Schmittlein, “Integration of the sales force: An empirical examination”, *Rand Journal of Economics*, Vol. 15, No. 3, pp385-395, 1984.
- [4A] Anderson & Weitz 1992 Anderson, Erin, and Weitz, Barton. “The Use of Pledges to Build and Sustain Commitment in Distribution Channels” *Journal of Marketing Research*. Vol 29, Iss. 1. p18-34.
- [5] Aubert, B. A., “Development of measures to assess dimensions of IS operation transactions”, *Omega*, Dec., pp13-25, 1996.
- [5A] Balakrishnan, Srinivasan, and Birger Wernerfelt. “Technical Change, Competition, and Vertical Integration” *Strategic Management Journal*. Jul/Aug 1986. Vol 7, Iss. 4. p347-359.
- [6] Brack, K., “Your e-options”, *Industrial Distribution*, Vol. 89, No. 7, Jul, pp54-58, 2000.
- [7] Bradach, J. L., R. G. Eccles, “Price, authority, and trust: From ideal types to plural forms”, *Annual Review of Sociology*, Vol. 15, pp97-118, 1989.
- [8] Bucklin, L., S. Sengupta, “Organizing successful co-marketing alliances”, *Journal of Marketing*, Vol. 57, April, pp32-46, 1993.
- [9] Chandler, A. D., *The Visible Hand: The Managerial Revolution in American Business*, Cambridge, MA: Belknap Press, 1977.
- [10] Clemons, E. K., M. C. Row, “Information, power, and control of the distribution channel” *Chief Executive*. New York: May 1993. p. 64 -48.
- [11] Coase, R. H. “The nature of the firm”, *Economica*, Vol. 4, November, pp386-405, 1937.
- [12] Daft, R. L., *Organization Theory and Design*, 5th ed., St, Paul, MN: West Publishing Company, 1995.
- [13] Earl, P. E., *Information, Opportunism and Economic Coordination*, Northampton, MA: Edward Elgar, 2002.

- [14] Eisenhardt, K. M., "Speed and strategic choice: How managers accelerate decision making", *California Management Review*, Berkeley, Vol. 32, No. 3, Spring, p39, 1990.
- [15] Elfenbein, D. W., J. Lerner, "Ownership and control rights in Internet portal alliances, 1995-1999", *The Rand Journal of Economics*, Vol. 34, No. 2, Summer, pp356-370, 2003.
- [15A] Erramilli, M. Krishna, and Rao, C.P. "Service Firms' International Entry-Mode Choice: A Modified Transaction-Cost Analysis Approach" *Journal of Marketing*. July 1993. Vol 57. p19-38.
- [16A] Nicolai Foss and Volker Mahnke., *Competence, governance, and entrepreneurship : advances in economic strategy research*, New York: Oxford University Press, 2000.
- [16] Hair, J. F., R. E. Anderson, R. L. Tathan, W. C. Black, *Multivariate Data Analysis* 5th ed., Upper Saddle river, NJ: Prentice Hall, 1998.
- [17] Heide, J. B., G. John, "Alliances in industrial purchasing: The determinants of joint action in buyer-supplier relationships", *Journal of Marketing Research*, Vol. 27, Feb., pp24-36, 1990.
- [18] Heide, J. B., A. S. Miner, "The shadow of the future: Effects of Anticipated Interaction and Frequency of Contact on Buyer-seller Cooperation", *Academy of Management Journal*, Vol. 35, No. 2, June, pp265-292, 1992.
- [19] John, G., "An empirical investigation of some antecedents of opportunism in a marketing channel", *Journal of Marketing Research*, Vol. 21, August, pp278-289, 1984.
- [20] John, G., T. Reve, "The reliability and validity of key informant data from dyadic relationships in marketing channels", *JMR, Journal of Marketing Research*, Vol. 19, No. 4, Nov, pp 517-525, 1982.
- [20-A] John, George, and Weitz, Barton. "Salesforce Compensation: An Empirical Investigation of Factors Related to Use of Salary Versus Incentive Compensation" *Journal of Marketing Research*. Feb. 1989. Vol 26. p1-14.
- [20A] Joskow, Paul L. "Contract Duration and Relationship-Specific Investments: Empirical Evidence from Coal Markets" *The American Economic Review*. March 1987. Vol. 77, Iss. 1. p168-185.
- [20-B] Kaufman, Patrick J. and Stern Louis W. "Relational Exchange Norms, Perceptions of Unfairness, and Retained Hostility in Commercial Litigation" *Journal of Conflict Resolution*. Sept 1988. Vol 32, Iss. 3. p.534-552
- [20B] Klein, Saul. "A Transaction Cost Explanation of Vertical Control in International Markets", *Academy of Marketing Science Journal*, Summer 1989. Vol. 17, Iss. 3; p. 253 -260.
- [21] Klein, S., V. J. Roth, "Determinants of export channel structure: The effects of experience and psychic distance reconsidered", *International Marketing Review*, Vol. 7, No. 5, pp27-38, 1990.
- [22] Klein, S., G. L. Frazier, V. J. Roth, "A transaction cost analysis model of channel integration in international markets", *Journal of Marketing Research*, Vol. 27, No. 2, May, pp196-208, 1990.
- [23] Langlois, R. N., P. L. Robert, "Networks and innovation in a modular system: Lessons from the microcomputer and stereo component industries", *Research Policy*, Vol. 1, No. 1, pp. 297-313, 1992.
- [23-A] Levy, David T. "The Transactions Cost Approach to Vertical Integration: An Empirical Examination" *The Review of Economics and Statistics*. Aug. 1985. Vol 67, Iss. 3. p438-445.
- [23A] Leffler, Keith B., and Rucker, Randal R. "Transactions Costs and the Efficient Organization of Production: A Study of Timber-Harvesting Contracts" *Journal of Political Economy*. 1991. Vol 99, Iss. 5. p1060-1087.
- [23B] Lieberman, Marvin B. "Determinants of Vertical Integration: An Empirical Test" *The Journal of Industrial Economics*. Sept. 1991. Vol 39, Iss. 5. p451-466.
- [24] Maltz, Arnold. "Outsourcing the Warehousing Function: Economic and Strategic Considerations" *Logistics and Transportation Review*. Sept. 1994. Vol 30, Iss. 3. p245-265.
- [25] Maltz, Arnold. "Private Fleet Use: A Transaction Cost Model" *Transportation Journal*. 1993. Vol 32, Iss. 3. p46-53.
- [26] Masten, S. E., "The organization of production: Evidence from the aerospace industry", *Journal of Law and Economics*, Vol. 27, No. 2, pp403-418, 1984.
- [27] Monteverde, K., D. J. Teece, "Supplier switching costs and vertical integration in the automobile industry", *Bell Journal of Economics*, Vol. 13, No. 1, pp206-214, 1982.
- [28] Nelson, R. E., "The strength of strong ties: social networks and intergroup conflict in organizations", *Academy of Management Journal*, Vol. 32, No. 2, June, pp377-401, 1989.
- [29] Noordewier, T. G., G. John, J. R. Nevin, "Performance outcomes of purchasing arrangements in industrial buyer-vendor relationships", *Journal of Marketing*, Vol. 54, No. 4, Oct., p80-93, 1990.
- [29A] Osborn, Richard N., C. Christopher Baughn. "Forms of Interorganizational Governance for Multinational Alliances" *Academy of Management Journal*. 1990. Vol 33, Iss. 3. p503-519.
- [30] Ouchi, W. G., "Markets, bureaucracies, and clans", *Administrative Science Quarterly*, Vol. 25, March, pp29-145, 1980.
- [31] Palay, Thomas M. "Comparative institutional economics: the governance of rail freight contracting", *the Journal of Legal Studies*, Vol. 13, No. 2, pp265-287, 1984.
- [32] Parkhe, A., "Strategic alliance structuring: A game theoretic and transaction cost examination of interfirm cooperation", *Academy of Management Journal*, Vol. 36, No. 4, pp794-829, 1993.
- [33] Pillin, B. K., L. A. Crosby, D. W. Jackson, Jr., "Relational bonds in industrial exchange: An experimental test of the transaction cost economic

- framework”, *Journal of Business Research*, Vol. 30, pp237-251, 1994.
- [34] Pinckney, B. “New law helps banks break barriers”, *Capital District Business Review*, Vol. 27, No. 3, May 1, p20, 2000.
- [35] Powell, W. W., “Neither market nor hierarchy: Network forms of organization”, *Research in Organizational Behavior*, Vol. 12, pp295-336, 1990.
- [36] Pugh, D. S., “The measurement of organizational structures. Does context determine form?”, *Organizational Dynamics*, Vol.1, No. 4, pp19-34, 1993.
- [37] Pugh, S., D. J. Hickson, C. R. Hinings, C. Turner, “Dimensions of organization structure”, *Administrative Science Quarterly*, Vol. 13, No.1, pp65-105, 1968.
- [38] Ring, P. S., A. H. Van de Ven, “Formal and informal dimensions of transactions”, in *Research on the management of innovation: The Minnesota studies*. Van de Ven, A. H., H. L. Angle, M. S. Poole, (editors), pp171-192, 2000.
- [39] Ring, P. S., A. H. Van de Ven, “Structuring Cooperative Relationships Between Organizations”, *Strategic Management Journal*, Vol. 13, No. 7, Oct., pp483-498, 1992.
- [40] Smith, A., *The wealth of nations*. London: W. Strahau & T. Cadell, 1776.
- [40-A] Spekman, Robert E., Salmond, Deborah J., Lambe, C. Jay., “Consensus and collaboration: norm-regulated behaviour in industrial marketing relationships” *European Journal of Marketing*. [Online]. Available: <http://proquest.umi.com/pqdweb?RQT=318&pmid=14913&TS=1095790099&clientId=30291&VType=PQD&VName=PQD&VInst=PROD.1997> <http://proquest.umi.com/pqdweb?RQT=572&VType=PQD&VName=PQD&VInst=PROD&pmid=14913&pcid=1656381&SrchMode=3>. Vol.31, Iss. 11/12; pg. 832.
- [40A] Sriram, Ven, Krapfel, Robert, and Spekman, Robert. “Antecedents to Buyer-Seller Collaboration: An Analysis From the Buyer’s Perspective” *Journal of Business Research*. 1992. Vol 25. p 303-320.
- [41] Stump & Heide, 1996
- [41A] Trabold, Harald. “Export intermediation: An empirical test of Peng and Ilinitch” *Journal of International Business Studies*, Washington: Second Quarter 2002; 33:2; pg. 327-345.
- [42] Walker, G., L. Poppo, “Profit centers, single-source suppliers, and transaction costs”, *Administrative Science Quarterly*, Vol. 36, No. 1, pp66-88, 1991.
- [43] Walker, G., D. Weber, “Supplier competition, uncertainty, and make-or-buy decisions”, *Academy of Journal Management*, Vol. 30, No. 3, Sept., pp589-596, 1987.
- [44] Walker, G., D. Weber, “A transaction cost approach to make-or-buy decisions”, *Administrative Science Quarterly*, Vol. 29, No. 3, pp373-392, 1984.
- [44A] Weiss, Allen M., and Anderson, Erin. “Converting from Independent to Employee Salesforces: The Role of Perceived Switching Costs” *Journal of Marketing Research*. Feb 1992. Vol 29, Iss. 1. p101-115.
- [45] Williamson, O. E., “Comparative economic organization: The analysis of discrete structural alternatives”, *Administrative Science Quarterly*, Vol. 36, No. 6, pp 269-296, 1991.
- [46] Williamson, O. E. “The Economics of organization: The transaction cost approach”, *American Journal of Sociology*, Vol. 87, No. 3, pp549-577, 1981.
- [47] Williamson, O. E., “Transaction cost economics: The governance of contractual relations”, *Journal of Law and Economics*, Vol. 22, No. 10, pp233-261, 1979.

A Study on Design Process Integration between Marketing and R&D of Notebook

Linglang Tang, Yubin Chiu, Hungsheng Lin

Graduate School of Management Yuan Ze University Taiwan, China
balltang@saturn.yzu.edu.tw

ABSTRACT

This study focuses on building a framework which calculates the weights of FAHP (Fuzzy Analytical Hierarchical Process) method, then integrates the analysis tools such as FQFD (Fuzzy Quality Function Development) and FFMEA (Fuzzy Failure Mode and Effects Analysis) into the framework to construct a two-phase product specifications evaluation process. Therefore, this study builds a framework and process of the new product specification evaluation which integrates the marketing attributes (the evaluation of the customers' demands), research attributes and development attributes and manufacture attributes (the evaluation of the failures and defects of the product specifications). It allows the product specifications which are produced by brainstorming to be proceeded the evaluation before the prototype test phase and find the optimum product specification. It also can further the companies to optimize the organization resources. This research focuses on not only determining, but also transferring the market attributes to the product R&D specification in order to realize the relationship between the consumers' demands and the product specification. According to the methodologies represented, this study attempts to fill the gap in the literature by providing an integrative research framework and offers this framework to contribute the tasks and operations in the initial phase of the new product development and build the framework of multidimensional product specifications evaluation. Finally, an example of Notebook is used to illustrate the proposed approach.

Keywords: Marketing attributes, Fuzzy theory, AHP (Analytical Hierarchical Process), QFD (Quality Function Development), FMEA (Failure Mode and Effects Analysis), FQFD (Fuzzy Quality Function Development)

1. INTRODUCTION

Due to the increasing and more diversification of the consumers' demands for Notebook product, it offers an opportunity which the OEM/ODM Notebook companies transfer to OBM (Own Branding & Manufacturing) companies through building own brand. Besides, it also can develop new business model and business profits. The related literatures about the product design concepts which focus on the demands of consumer are always refer to QFD (Quality Function Deployment). QFD is a structured approach to ensure customers' demands to be satisfied and transform into the product design and development process. And it can contribute to increase in customer satisfaction and shorten product design and development time. However, existing QFD implementations have limitations. Two main problems are summarized as follows:

1. QFD unable to evaluate and integrate the possible problems in advance during the design and manufacture period of product development.
2. In the evaluation process, experts often exhibit some forms which are often vagueness and uncertainty and hinder the experts' description of conceptual phenomena.

For this reasons, the main objective of this paper is to build a framework and process of the new product specification evaluation which integrates the marketing attributes (the evaluation of the customers' demands), research and development attributes and manufacture attributes (the evaluation of the failures and defects of

the product specifications). It can allow the product specifications which are produced by brainstorming to be proceeded the evaluation before the prototype test phase and find the optimum product specification. It also can further the companies to optimize the organization resources.

Through the framework of two-phase product specifications evaluation is represented. Firstly, it performs FAHP (Fuzzy Analytical Hierarchical Process) to collect marketing experts with regard to authority evaluation of demands of customer and integrates it into FQFD (Fuzzy Quality Function Development) to proceed first phase evaluation of 「the degree of the consumers' satisfaction」. Secondly, to collect the evaluation of FFMEA (Fuzzy Failure Mode and Effects Analysis) of the design and manufacture experts about the product specifications which is allowed into the second phase and integrate it into FQFD to proceed second phase evaluation of 「the degree of the assembly manufactured」. Finally, it is able to find the optimum product specifications and allow it into the prototype test phase.

According to the methodologies represented by the paper, this study attempts to fill the gap in the literature by providing an integrative research framework and offers this framework to contribute the tasks and operations in the initial phase of the new product R&D and build the framework of multidimensional product specifications evaluation. Finally, an example of

Notebook is used to illustrate the proposed approach.

2. LITERATURE REVIEW

2.1 The marketing attributes

Sung et al.[10] showed 48 dimensions of usability were identified, which were classified into two groups based on the new definition of usability. The first group includes performance dimensions that could be used to explain the performance side of the usability concept. The second one consists of image/impression dimensions that are related to the image and impression of the product perceived by the users. Besides, aesthetic attributes are evaluated concurrently with analytically obtained objective attributes of the product's characteristics. The interaction of these aesthetic and objective attributes, as it relates to the product's utility and manufacturing cost, is clarified [14]. Sethi [8] indicated new product quality has been found to have a major influence on the market success and profitability of a new product. His research founded that quality is positively influence on the product development process, and quality orientation in the firm. But, moreover, the results showed product innovativeness has a negative effect on quality. And functional diversity and time pressure have no effect on new product quality. The relationship between information integration and new product quality is weakened by quality orientation in the firm. Information integration mitigates the negative effect of product innovativeness on quality.

Based on the above, this study is purported to classify and form the marketing attributes of Notebook. Consequently, the definition of the marketing attributes for Notebook product will be proposed. It is composed of the quality of product performance, the dimension of image/impression, and the the dimension of price. The three dimensions will be discussed later.

2.2 AHP and FAHP

Saaty [7] showed the Analytic Hierarchy Process (AHP) that is known as an additive pair wise weights identification method. AHP is often applied to find the relative weighting of each criterion.

Nevertheless, there are many researches, which pointed out that the assumption of AHP may cause the evaluation results departure from the experts' opinions. Therefore, there are many researchers using FAHP (Fuzzy Analytical Hierarchical Process) to resolve the problem. Laarhoved and Pedrycz [5] used Fuzzy Sets Theory and Fuzzy Arithmetic. Buckley, Feuring and Hayashi [1] indicated that a new method of finding the fuzzy weights in fuzzy hierarchical analysis which is the direct fuzzification of the original method used by Saaty[7] in the analytic hierarchy process is presented. Fuzzy number is introduced in the pairwise comparison of AHP by Kwong and Bai [4]. An AHP based on fuzzy

scales is proposed to determine the importance weights of customer requirements. They founded the new approach can improve the imprecise ranking of customer requirements which is based on the conventional AHP.

2.3 Failure Mode and Effects Analysis, FMEA

Thomas [11] pointed out that FMEA method could examine all potential causes or modes of failure, of critical processes and of methods designed to prevent failure of those processes. That is, FMEA examines all potential causes or modes of failure, of critical processes and of methods designed to prevent failure of those processes. Sharon [9] showed how key aspects of quality function deployment (QFD) and failure mode and effects analysis (FMEA) can be used in product and service development at a strategic level rather than in traditional engineering applications.

Pillay and Wang [6] evaluated the risk analysis tool FMEA assumes a failure mode, which occurs in a system/component through some failure mechanism; the effect of this failure is then evaluated. They described a risk ranking is produced in order to prioritize the attention for each of the failure modes identified. The traditional method utilizes the risk priority number (RPN) ranking system. This method determines the RPN by finding the multiplication of factor scores.

2.4 Fuzzy QFD

Yang et al.[13] presented the findings of a research effort to adapt House of Quality (HOQ) to meet the needs of buildable designs in the construction industry and to develop a fuzzy QFD system. They thought the differences between the proposed fuzzy QFD system and the traditional QFD methodology is that the QFD-relevant data are expressed and represented as linguistic terms rather than as crisp numbers, and the linguistic data is processed by algorithms embedded in the system's internal environment [13]. Chen et al. [2] determined the revised priority of the customer demands using a fuzzy logic inference because many tools did not offer specific methods to determine a revised priority for product redesign.

We transfer the market attributes to the product R&D specification in order to realize the relationship between the consumers' demands and the product specification.

3. METHODOLOGY

3.1 Marketing attributes of Notebook

Based on the literature review, we provide a definition of the marketing attribute for Notebook product, which is composed of the quality of product performance, image/impression, and price.

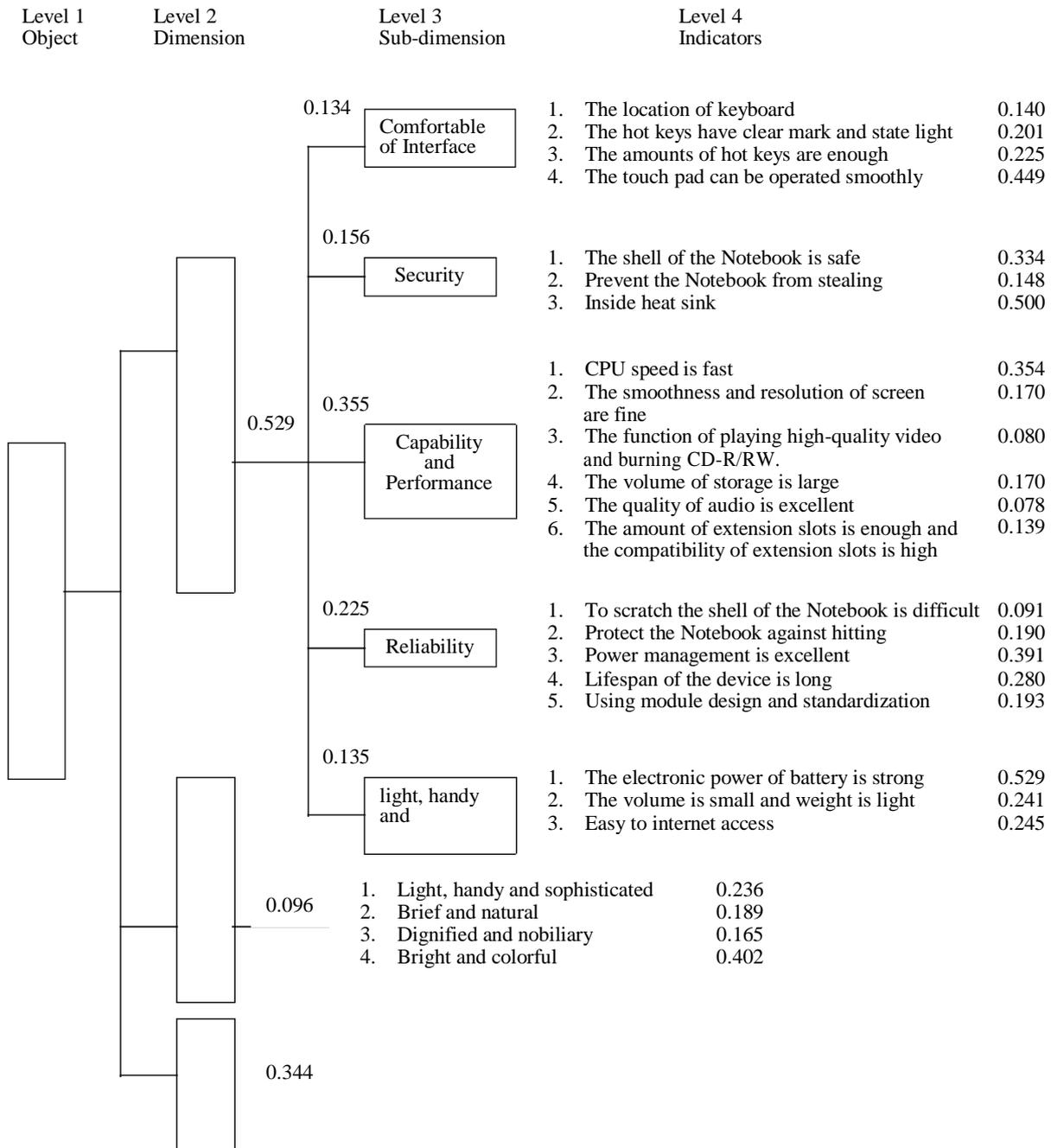


Figure 1 The weights of the marketing attributes

3.1.1 The quality of product performance

According to the researches made by Garvin[3] and Sethi[8]? Sung *et al.*[10] we defines the quality of product performances such as ? comfortable of interface?, ? security?, ? capability and performance? , 「reliability」 and? light, handy and convenient? .

3.1.2 The dimension of image/impression

Yamakawa *et al.* [12] indicated aesthetic attributes are very important while designing new product except high product performance and quality. In order to develop the items of image/impression

dimension in the Notebook market, study results from related areas were examined. This study follows the example of Sung *et al.*[10] to proceed extensive literature survey and find out the similes or metaphoric expressions to describe the image of the consumer Notebook products and various expressions about subjective feelings toward them. Also the experts and consumers opinions were collected, furthermore, the principal concepts used in the product design department were studied. Finally, we define the dimension of image/impression for Notebook products.

3.2 The integration of analysis methods

In order to complete the new product specification

evaluation, firstly, we use FAHP method to collect marketing experts with regard to authority evaluation of demands of customers, and integrates it into FQFD to proceed first phase evaluation of 「the degree of the consumers' satisfaction」. Secondly, to collect the evaluation of FFMEA of the design and manufacture experts about the product specifications and integrate it into FQFD to proceed second phase evaluation of 「the degree of the assembly manufactured」. Thus, we could find the optimum product specifications and allow it into the prototype test phase through the framework of two-phase product specifications evaluation.

This study defines the evaluation rules of consumer satisfaction and assembly manufactured scores. The first phase evaluation of 「the degree of the consumers' satisfaction」 would be accomplished. The score of consumers' satisfaction in the first stage should not be under five points, otherwise it will be eliminated or corrected its specification. The equations (1) and (2) are listed below.

$$CS_i^k = \left(\sum_{j=1}^n \overline{E}_j^k \times \overline{R}_{ij} \right) \times C_i \quad 1 \leq i \leq m \quad (1)$$

$$CS^k = \sum_{i=1}^m CS_i^k \quad 1 \leq i \leq m \quad (2)$$

CS_i^k : the i item satisfaction of product k

\overline{E}_j^k : the j item specification degree of product k

\overline{R}_{ij} : standardization triangular fuzzy membership number of the i marketing attribute item and j specification item

C_i : the weights of the i marketing attribute item

CS_i^k : the total satisfaction of product k

According to the data evaluated at the first phase, we could proceed the second phase, named evaluation of 「the degree of the assembly manufactured」. IF the result is not under five points, the product prototype can proceed pilot runs, otherwise it will be eliminated or corrected its specification. The equations needed (3) and (4) are listed below.

$$AD_i^k = \left(\sum_{j=1}^n \overline{E}_j^k \times \overline{R}_{ij} \times \overline{F}_j^k \right) \times C_i \quad 1 \leq i \leq m \quad (3)$$

$$AD^k = \sum_{i=1}^m AD_i^k \quad (4)$$

AD_i^k : the i item assembly manufactured degree of product k

AD^k : the total assembly manufactured degree of product k

4. EXAMPLE OF NEW PRODUCT SPECIFICATION EVALUATION

4.1 The weights of FAHP analysis for the marketing attributes

This study found six marketing experts to proceed the FAHP analysis about the multimedia entertainment style Notebook. The result is showed as Figure 1 (See Figure 1 for the weights of the marketing attributes). It indicates the most important dimension is the quality of product performance (0.529). In the next place, the price (0.344) dimension will be considered. The last one is 0.344 for the image/impression dimension. Furthermore, the sub-dimension named Capability and Performance is the most important, it got 0.355 points. So we may see the development direction in the future from the data.

4.2 The FQFD analysis for the marketing attributes and product R&D specifications

In order to transfer the market attributes to the product R&D specification, the FQFD method was proceeded. Six professional specialist in the R&D and assembly manufactured area participated the analysis process.

4.3 The first phase evaluation of 「the degree of the consumers' satisfaction」

Through survey of Notebook consumers, we got the comparison result of consumers' satisfaction and assembly manufactured between two kinds of products, and the table is presented as follows (See Table 1 for result). According to the data, the satisfaction score of product A is 5.598 and B is 5.682, so, these two kinds of products could proceed the second phase, the evaluation of 「the degree of the assembly manufactured」 because the satisfaction scores both these two products get are more than 5 points.

4.4 The possible problems evaluation in R&D and assembly manufactured

To find out the possible problems of these two kinds of products, six professional specialists in the R&D and assembly manufactured area participated the FFMEA analysis process. They transferred the possible problems in different conditions to the FRPN (Fuzzy Risk Priority Number). The FRPN value is between 1 to 9. The higher the score, the more serious the problem. And table 2 illustrates the weighting of problems in two kinds of products.

4.5 The second phase evaluation of 「the degree of the assembly manufactured」

Finally, we integrate the FRPN of each possible problem into the specification to proceed second phase evaluation of 「the degree of the assembly manufactured」. As table 1 shows, the degree of the assembly manufactured for product A is 4.964 and B is 5.135. It illustrates that product A should be corrected in

some specifications, even its satisfaction score is 5.598 there are still problems. Product B could proceed pilot runs. On the other hand, the total gap of satisfaction between A and B is 0.085 (1.5%), but the total gap of Assembly manufactured between A and B is 0.172 (3.5%). It conspicuously increases.

5. CONCLUSIONS AND SUGGESTIONS

The conclusions of this study are as follows: According to the methodologies represented, this study attempts to fill the gap in the literature by providing an integrative research framework and offers in this framework to contribute the tasks and operations in the initial phase of the new product development and build the framework of multidimensional product specifications evaluation. Besides, we provide a methodology to integrate the FAHP, FQFD and FFMEA methods. So, through considering the consumers' demands and the new product specification evaluation, companies could integrate the opinions of different departments.

This research focuses on not only determining, but also transferring the market attributes to the product R&D specification in order to realize the relationship between the consumers' demands and the product specification so that it can further the companies to optimize the organization resources.

REFERENCES

- [1] Buckley, J. J. and Feuring, T. and Hayashi, Y., "Fuzzy hierarchical analysis revisited", *European Journal of Operational Research*, Vol.129, pp.48-64, 2001.
- [2] Chen C. K., Chen L. C. and Lin L., "Methods for processing and prioritizing customer demands in variant product design", *IIE Transactions*, Vol.36, Iss. 3; pg. 203, 2004.
- [3] Garvin, D. A., *Managing Quality: the Strategic and Competitive Edge*, New York: Free Press, 1988.
- [4] Kwong, C.K. and Bai, H., "A fuzzy AHP approach to the determination of importance weights of customer requirements in quality function development", *Journal of Intelligent Manufacturing*, Vol.13, pp.367-377, 2002.
- [5] Laarhoven, P.J.M. & Pedrycz, W., "A fuzzy extension of Saaty's priority theory", *Fuzzy Sets and Systems*, Vol.11, No.3, pp.229-241, 1983.
- [6] Pillay A. and Wang J., "Modified failure mode and effects analysis using approximate reasoning", *Reliability Engineering and System Safety*, Vol.79, Iss.1, pp.69-85, 2003.
- [7] Saaty, T. L., *The Analytic Hierarchy Process*, McGraw-Hill, New York, 1980.
- [8] Sethi, R., "New product quality and product development teams", *Journal of Marketing*, Vol.64, Iss.2, pp.1-15, 2000.
- [9] Sharon K Johnson, "Combining QFD and FMEA to optimize performance", *Quality Congress. ASQ's ... Annual Quality Congress Proceedings*, Milwaukee, pp. 564-575, 1998.
- [10] Sung, H.H. and Yun, M. H. and Kwahk, J. and Hong, S.W., "Usability of consumer electronic products", *International Journal of Industrial Ergonomics*, Vol.28, pp.143-151, 2001.
- [11] Thomas T Reiley, "FMEA in preventing medical accidents", *Quality Congress. ASQ's ... Annual Quality Congress Proceedings*, Milwaukee, p. 657-664, 2002.
- [12] Yamakawa, H. and Katsutyama, N.O.M. and Hiroyasu, T., "A study on heredity and evolution of designs considering Kansei by using genetic algorithms", *Proceedings of International Symposium on Optimization and Innovative Design*, Vol.194, pp.1-9, 1997.
- [13] Yang, Y.Q. and Wang, S.Q. and Dulaimi, M. and Low, S.P., "A fuzzy quality function deployment system for buildable design decision-makings", *Automation in Construction*, Vol.12, pp.381-393, 2003.
- [14] Yoshimura, M. and Yanagi, H., "Strategies for implementing aesthetic factors in product designs", *International Journal of Production Research*, Vol.39, Iss.5, pp.1031-1049, 2001.

Table 1 The comparison of consumers' satisfaction and assembly manufactured between two kinds of products

Dimension	Sub-dimension	Consumers' satisfaction score		Assembly manufactured score		The gap of consumers' satisfaction	The gap of Assembly manufactured	Corrected value
		NB A	NB B	NB A	NB B			
Comfortable of interface	The location of keyboard	6.124	5.572	4.535	4.268	0.552	0.267	0.285
	The hot keys have clear mark and state light	6.136	5.791	5.133	4.874	0.345	0.259	0.086
	The amounts of hot keys are enough	6.242	5.693	5.338	4.807	0.549	0.531	0.018
	The touch pad can be operated smoothly	6.262	5.034	3.813	3.885	1.228	(0.072)	1.300
Security	The shell of the Notebook is safe	5.459	5.902	5.459	5.902	(0.443)	(0.443)	0.000

	Prevent the Notebook from stealing	5.366	5.542	5.366	5.542	(0.176)	(0.176)	0.000
	Inside heat sink	4.730	5.909	4.500	5.909	(1.179)	(1.409)	0.230
Capability and Performance	CPU speed is fast	4.441	6.510	4.235	6.282	(2.069)	(2.047)	(0.023)
	The smoothness and resolution of screen are fine	4.809	6.745	3.756	5.805	(1.936)	(2.049)	0.113
	The function of playing high-quality video/burning CD	5.167	5.674	3.420	1.962	(0.507)	1.459	(1.965)
	The volume of storage is large	5.343	6.368	2.312	5.324	(1.025)	(3.012)	1.987
	The quality of audio is fine	3.636	6.457	2.234	4.214	(2.821)	(1.980)	(0.840)
	The amount of extension slots is enough and the compatibility of extension slots is high	5.082	6.070	3.905	4.288	(0.988)	(0.384)	(0.604)
Reliability	To scratch the shell of the Notebook is difficult	5.389	5.926	5.389	5.926	(0.537)	(0.537)	0.000
	Protect the Notebook against hitting	5.523	5.919	4.779	5.259	(0.396)	(0.479)	0.083
	Power management is fine	4.449	5.869	2.949	4.932	(1.420)	(1.982)	0.563
	Lifespan of the device is long	5.408	5.643	5.408	5.643	(0.235)	(0.235)	0.000
	Using module design and standardization	5.805	5.905	5.805	5.905	(0.100)	(0.100)	0.000
light, handy and convenient	The electronic power of battery is strong	4.507	5.671	2.868	4.489	(1.164)	(1.621)	0.456
	The volume is small and weight is light	5.129	5.748	3.953	4.605	(0.618)	(0.653)	0.034
	Easy to internet access	5.545	5.693	5.545	3.306	(0.148)	2.239	(2.387)
image/impression	5.834	5.645	4.706	4.723	0.189	(0.017)	0.206	
price	6.448	5.403	6.448	5.403	1.045	1.045	0.000	
Total of score	5.598	5.682	4.964	5.135				
Total of gaps					(0.085)	(0.172)	0.087	
Percentage of gaps					(1.5 %)	(3.5 %)		

Table 2 The weighting of problems in two kinds of products

Problem weighting		Product A		Product B	
		R&D problem weighting	Assembly manufactured problem weighting	R&D problem weighting	Assembly manufactured problem weighting
Display module	LCD specification	35.318	33.744	5.760	26.416
	LED	5.7605	0	5.760	0
Operation interface	Keyboard device	5.7605	5.480	8.718	5.4780
	Touch pad device	13.703	0	0	5.949
Storage device	Hard disk device	9.923	8.251	0	0
	CD-ROM device	2.962	2.211	20.992	7.964
Motherboard	Extension slots	5.949	12.975	9.689	12.975
	System BIOS	4.453	4.453	4.453	4.453
	Power management	5.949	5.480	0	0
	Battery	26.228	12.859	13.307	6.653
Device	Sound card module	7.412	7.412	5.949	5.4780
	Network device	0	0	11.899	10.401

A Web-based Operation Management System for Distributed Divisional Organizations

Wei Shang, Yijun Li, Qiang Ye, Wenjun Sun

School of Management, Harbin Institute of Technology, Harbin 150001, China
{shangwei, liyijun, yeqiang, wjsun}@mail.hit.edu.cn

ABSTRACT

Operation Management is an important and complex task for a divisional structured organization, especially when the divisions are distributed geographically. In most cases, such organizations didn't not urge all of it's divisions to use an integrated information system at the very beginning. But with the development and the expanding of the organization, they sometimes found themselves in the trouble of information exchange and almost lost control of their divisions. At such time, however, on one hand the head quarter inquires more detailed information and more business control on the divisions. On the other hand, some divisions are well built and have its own business processes and information systems. It's impossible for them to rebuild the information system to integrate with the other divisions and the head quarter as well. Operation Management System (OPMGT) enables real-time inspection of the divisions' operational data and flexible operation evaluation of each division via the Internet and without much change on the other information systems. The OPMGT presented in this paper was originally developed for the head quarter of a distributed divisional based organization to govern the distributed divisions. System analysis, design and implementation of OPMGT are discussed in detail. Having been developed on the basis of eFramework, a J2EE framework, OPMGT is proved to be highly sufficient in operation management of a distributed divisional structured organization, and it may also do some help to integrate information systems in some degree.

Keywords: Operation Management System, Distributed divisions, Enterprise Information System, J2EE framework, Operation Evaluation

1. BACKGROUND

With the development of information technologies, Enterprise Information System (EIS for short) has been playing an important role in the enterprise management. The new generation of eBusiness (electronic business) becomes more rational and focuses more on the IT (Information technology) revolution inside the organizations [1]. And the enterprises tend to be geographically distributed with the development of the global economics. To control the operation of the whole organization, information system together with the Internet is a possible facility. There are three basic organization structures in practice: functional, divisional and matrix. Different kind of organization structure calls on different emphasis in operation management.

In a functional structure, each part of the organization should work together in a coordinated manner and operational management should be performed upon the vision of the whole organization. While in a divisional structured organization, each division is quite independent financially in most cases, and the top management usually does not interfere with the operation details of each division, but tends to supervise their operation situation as a whole. Matrix structured organization simultaneously implements both structures mentioned above by having a division manager and a functional manager at the same time to keep control of productive and managerial affairs.

The operation management function in most existing EIS is designed for functional structured organization. And in a matrix-structured organization, project management is the commonly used and effective method to manage the operation of the organization. As for the divisional organization, unlike the other two types, cooperation among divisions is loose and organizations today might have their divisions in different cities or even different countries. Unfortunately, in many geographically distributed organizations, operation management is usually emphasized inside each division only, and operation management of the whole organization is neglected. The importance of developing an effective operation management system for distributed divisional structured organizations has been recognized. We have built such a system, called OPMGT, which is web-based and tailored for a design company of the petroleum industry to manage their distributed divisions.

System requirements are discussed in the following section. The eFramework on which we built our OPMGT is introduced in the third section. Then there goes the detailed design and development. Implementation and conclusions are discussed as the last part of the paper.

2. SYSTEM REQUIREMENTS

OPMGT is an information system for operation management, especially for distributed divisional structured organizations. Here, we limit our research in

the organizations that have more than three divisions and tend to control the operation of each division at a relatively high level instead of a detailed level. Since the divisions might be distributed geographically, we should build our system as an Internet based application. Only in this way, different divisions can access the system easily and interact with the head quarters continently. To understand what problems the OPMGT solves and how the OPMGT works, we need to take a look into the operation system requirements and infrastructure.

To supervise the divisions and make a future strategy, the top executives of a divisional organization usually care about the following aspects of the divisions' operational affairs.

- How do they fulfill their part of the whole organization's goal? Do they make more profits or do they spend more over the annual budget? Finding the weak points of the organization's operation is essential to improve the performance of the whole organization.

- How to evaluate their operation effects compared with other divisions? Bonus distribution among the divisions is not an easy task. You cannot just scatter the money according to the profit per person. Some divisions do not make much profit or even do not make any profit at all, but their contributions to the overall organization are the same or even more.

- How to make the anticipated profit or cost-based budget of the next year referenced with their historical operation data? Identifying the increase or decrease in efficiency is due to accident or an unavoidable trend is a rather statistical problem that needs the support of a large amount of operational data.

- How the division manager's leadership affects the whole division's operating situation? Since the division is independent in many aspects, the divisional manager is quite important to the performance of the whole division.

The employees in each division are also concerned with their performance in some of or all of the above aspects. The key of operation management in divisional structured organizations is how to evaluate each division. This measurement should be fair and open.

Technically, we adopt a Web-based J2EE framework to build the OPMGT as a browser and server (B/S) application. In this way, the system can be accessed anywhere through the Internet and has the enterprise features of expansibility and reusability.

3. A WEB-BASED J2EE FRAMEWORK

3.1 Infrastructure

The JAVA 2 Enterprise Edition (J2EE) is one of the most popular technologies used in EIS development by its features of platform-independent, object-oriented, component-based and multi-tier-supported. There are

more and more enterprise applications developed totally or partly using the J2EE technology. J2EE framework is a set of tools to enhance the efficiency and reliability of the enterprise application development. Here we introduce a web-based J2EE framework, eFramework (developed and maintained by Polysoft International Corp.), which is especially designed for EIS.

J2EE introduced a component-based method to design, develop and deploy enterprise applications. Despite of the various J2EE design patterns, J2EE infrastructure is composed of three tiers: client, server and background system (e.g. the database system). The J2EE specification defines the following component:

- Client component, which is a web page browser to parse the HTML document and support the JavaScript dynamic effects.

- Server component, which contains Servlets and JSPs.

- Business Component, which is possibly implemented by EJBs.

eFramework supports the new stable version of JSP and includes the newest Servlet APIs. EJB is included by eFramework as a solution for complex business logic. Applications based on eFramework can be deployed to any application server that supports the J2EE specifications.

3.2 The MVC design pattern in eFramework

The Model-View-Controller (MVC) design pattern has been recognized as a brilliant software design pattern that enforces the separation between the input, processing, and output of an application. eFramework adopts MVC pattern in web-based applications with J2EE technology. (Shown as Figure1.)

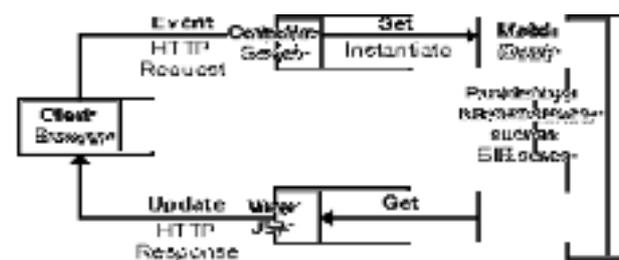


Figure1. M-V-C design pattern of eFramework

When a HTTP request reaches the web server, the following events occur:

- All the requests are captured by the Controller.

- Controller receives the request from a browser and decides where to send it by parameters assigned before hand.

- Controller sends instructions to Handler Factory and demands for a specific Handler to process the request.

- Controller commands the Handler to deal with the request.

- ó Handler presents the Model of M-V-C pattern, traces the states and implements the business logic. Each Handler knows its duty of combining multiple data beans into a Page Bean.

- ó Handler asks Persistence Manager for data and put these data into a web Page Bean.

- ó Persistence Manager is responsible for saving and getting data through JDBC and hides the detail of stateless session beans to implement the encapsulation rules of software engineering.

- ó If any exceptions occur during the process above, the Handler will decide whether to generate exception pages, for instance, error page beans specially designed for present error messages.

- ó Handler passes the assembled Page Bean to Controller.

- ó The Page Bean, which has been sent to the Controller, will give out its Presenter.

- ó The Controller passes the user request and response parameters to the Presenter, and asks them to present the Page Bean.

- ó The JSP page appointed in the Presenter only needs to get data from the Page Bean, convert them into HTML document and send it back to the user.

The whole closed loop is shown in Figure 2:

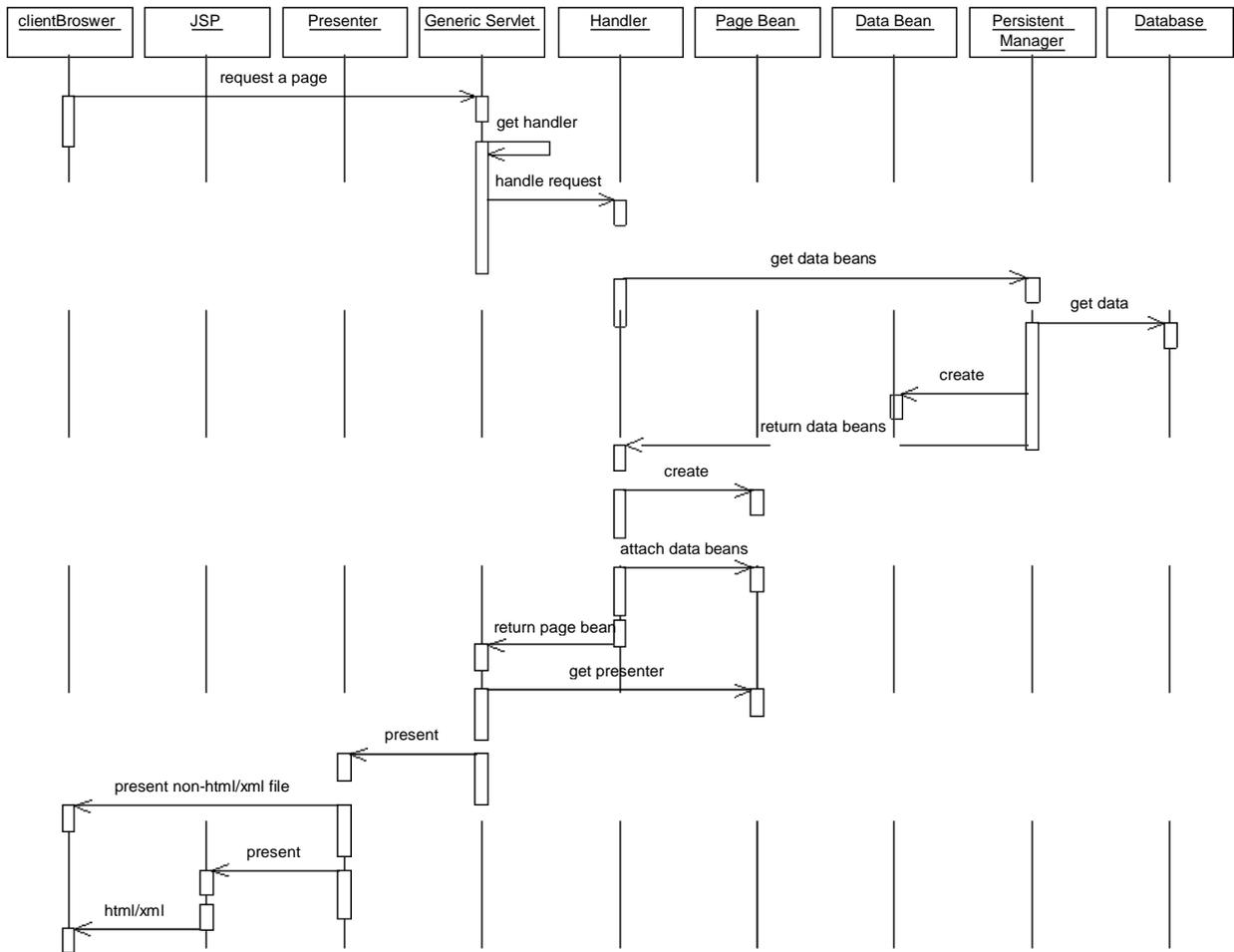


Figure2. M-V-C design pattern of eFramework

3.3 Features

Presentation, business logic and persistent data management are separated clearly in the eFramework. Multi-language, customizing, and application diagnostics are supported as well. By supporting the mechanism of independent development on JSPs and Servlets, eFramework also enables non-Java developer to built JSP pages. Multi-tiered logging functions

enhance the just-in-time inspection and effective debugging on the system. J2EE infrastructure enables flexible deployment on any Servlet container and JSP Engine compliant with the J2EE specifications. Although eFramework seems more complicated than other J2EE frameworks, such as struts and cocoon etc., it is more suitable for building large EIS for its outstanding features.

4. SYSTEM DESIGN

4.1 System structure

Main function of OPMGT is to help the executives take control of the operation management. OPMGT should

be a subset of an EIS, therefore coordination with other part of the EIS must also be carefully considered. The system structure is shown in Figure 3.

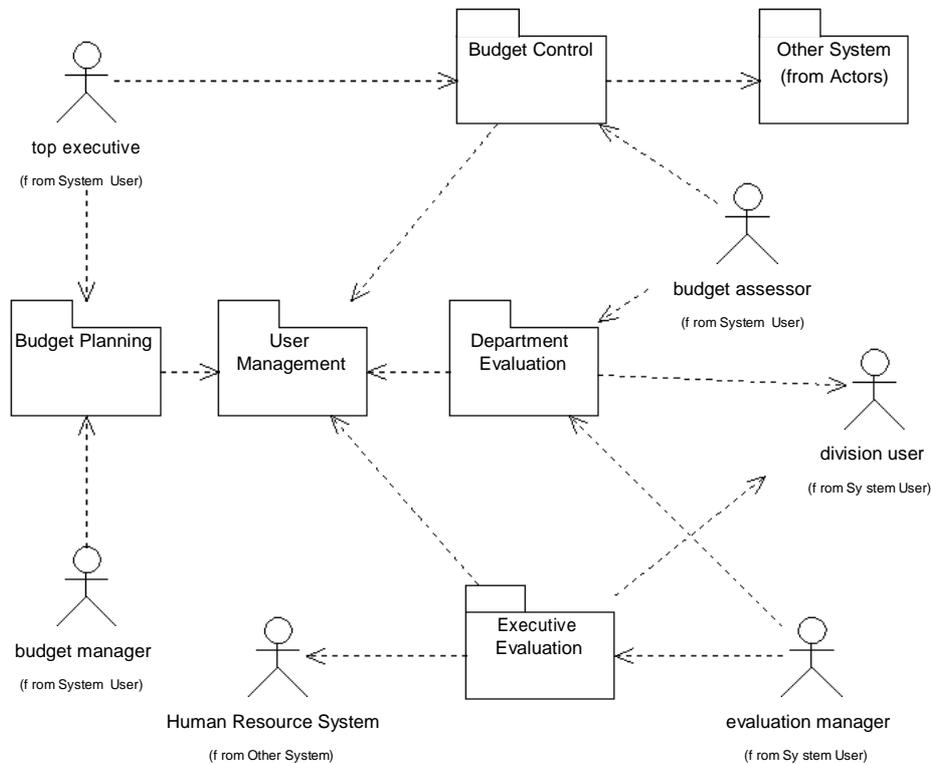


Figure 3. System Structure Diagram

Five packages are needed to satisfy the system requirements: user management, budget planning, budget control, department evaluation, and executive evaluation.

4.2 Database management strategy

eFramework provides a database connection pooling management mechanism to manage the JDBC database connections. Since OPMGT is related to several databases, several connection pools should be built for the system to contact with the background databases. Information of these connections is specified in a special type of configuration files (java property files to be specific). This kind of file may have its content changed without having to force rebuilding of the whole project. Therefore, agile development on database connection management is implemented.

Small data unit, such as user information, is designed as a Java class. A persistent manager is a final class to manage the instances of those classes that interact with the databases to obtain small data units. This helps to

encapsulate the database access by letting eFramework manage the lower level details of data exchange and ensure data transfer security and data consistency.

In OPMGT, historical data may be required for querying. We design a backup database to store a more recent historical data, for example, a two-year old dataset. When a request for historical data occurs, a connection to the backup database is built dynamically. After the processing of historical data, the connection is released automatically, since there is not a single active communication for a certain time. With the aid of eFramework, developers may concentrate on the database management logic and leave the routine database communication to eFramework.

4.3 User authentication and authorization design

User authentication and authorization is pretty indispensable in OPMGT. Without authentication, a user can only access very limit information, for most of the operation data of an organization could be confidential. Authentication is not enough, in the case that users of

different managerial levels need to have different levels of rights to use the system. A top executive checks raw operating data of each division but cannot modify it. While a common division user may evaluate the operation of his division but cannot change the evaluating method. Authorization is complex and may change often during the run-time of the system also.

We therefore design two inheritable handlers: `AuthenticatedRequestHandler` and `AuthorizedRequestHandler`. The former class contains the default user authentication logic that ensures that the user has been logged into the system with right user ID and password. The latter class checks if a user has the right to do a certain operation by checking whether the request identity sent by the user has been assigned legal for the user to execute. The relations between user IDs and corresponding request identities are stored in a table. Only the system administrator has the right to maintain this information.

By inheriting from the `AuthenticatedRequestHandler` or the `AuthorizedRequestHandler`, each handler has its own security level. The build-in rules of authentication and authorization release the developer of each module from the trifle work of deciding the user's rights in every operation.

4.4 Detailed design of the system

4.4.1 Budget planning module

The budget-planning module is concerned with the operation control in the organization level. Making the annual income budget and outcome budget is the main function of this module. Tracing the implementation of the budget and compare the actual budget with the planning budget is another function of this module. An annual budget may contain several sub budgets for each division. Both the top executives and the division managers have the rights to check these budgets.

Making and tracing these budgets with a visual statistic chart may help all level of the managers to have an impressive understanding of the overall situation of the organization. We design a statistic figure generation package based on the `JFreeChart` library. A chart factory manages the instance of three kinds of charts: pie chart, bar chart and line chart. When a user requests to draw a figure, the chart factory offers a specific chart drawing class and return the user a JPG format figure. This statistic figure generation package can be used in other part of the OPMGT also.

4.4.2 Operation data management

There are two means to gather the operation data: collecting from the existing information system and inputting it into the OPMGT manually. If there's already electronic data in the EIS, this kind of data could be

read from the enterprise database directly. But in some cases the data may be collected by commercial software that does not open its database. For example, the expense flow may have been collected by an existing accounting information system. In this instance, the OPMGT has to import some kind of output document from that software. The output document could be a Microsoft Excel file or a XML file or some other formatted data file. We design an XML import module based on Xalan APIs and an Excel import module based on JXL library to import existing operation data from existing EIS. If not all the operation data has been collected electronically, it's the OPMGT's responsibility to propose a convenient, user-friendly operation data input interface. An input form with combo boxes and check boxes is designed as a manually data input interface in OPMGT.

Querying and summarizing the operation data is also designed as an important function in operation data management module. Once the data has been input into the system, nobody can revise them or delete them. If there's any mistake in the raw input data, the only thing to correct it is to add a negative income or outcome noted the corresponding income or outcome ID. This rule just presents the common sense of accounting information system.

4.4.3 Division evaluation and executive evaluation

Evaluation measurement of the division's performance could be complicated and inconstant. Accordingly, we develop an equation parsing and calculating package. The equation is converted into a reverse Polish notation and a binary tree is used to present the equation and the sub equations. With the flexible equation mechanism the evaluation measurement can be defined and revised in the run time. Some variables in the evaluation equation are from the gathering of the operating data and others may be coefficients which value is different according to different year and different division. These two kind of variable are distinct in that gathering of the operating data is dynamic and could be obtained from the system automatically, while the value of the coefficients are decided subjectively by the evaluating operator. Thus we present and store these two kinds of variable separately.

The executive evaluation differs from the division evaluation only in the presentation of the equations and the source of the variable values. Therefore, we use the same module but different storage to design the executive evaluation module.

5. SYSTEM IMPLEMENTATION AND CONCLUSIONS

The OPMGT is developed among a small group consists five developers. We built a CVS depository to store the code and coordinate the team development process.

Taking the advantages of the eFramework, the development process is efficient and efficient. Testing and debugging work is also smooth according to the well-structured design pattern and the eFramework logging mechanisms. Since the OPMGT do not have heavy traffic, an Apache server with a tomcat plug-in is enough for a web application server.

In developing the OPMGT, eFramework is proved to be an efficient in information system development. But due to its powerful infrastructure, it is not easy for a developer to master the spirit and exert the best performance of the framework. The starting of our development is much harder than the following period. Consequently, a long term and relatively steady develop team may be better for building an information system using this eFramework.

The authentication and authorization mechanism is sufficient for the access control of the system functions. The flexible design of the evaluation equation enables various dynamic measurements of the divisions in an organization. The organization testing the OPMGT has found it practical and effectual in managing the operation of the divisions as a whole. Further more, as an application based on J2EE technology, the OPMGT has the cross-platform feature and this makes it possible to be integrated into existing enterprise system. The excel file input and figure output functions also make it

convenient to exchange data and express information in a distributed divisional structured organization. The design company who raised the request of building this system is very satisfied with it. However, they have not used it for long time. And with time past, more improvement of the system must be done.

ACKNOWLEDGEMENT

This research is sponsored by the Natural Science Foundation of Heilongjiang, PRC (G0304) and the Key Sci-tech Project Plan of Harbin, PRC 2004AC9CT073).

REFERENCES

- [1] Erik Brynjolfsson, Glen L. Urban, *Strategies for e-business success*, MIT Sloan Management Review, Nov. 2001, pp, 38-43
- [2] James I. Cash, Building the *Information-Age Organizations: Structure, Control, and Information Technologies*, McGraw-Hill Book Co. 1998, pp, 28-32
- [3] Deepak Alur, Jhon Crupi, Dan Malks. "Sun Java Center J2EE Patterns", <http://developer.java.sun.com/developer/technicalArticles/J2EE/patterns/>, 2002.5.14
- [4] Yan Hong, *Java and Patterns*, Publishing House of Electronics Industry, Beijing,

Apply MDA to E-Business: MDA Based Workflow Solution as an Example

Jiawei Lee, Jungsing Jwo

Department of Computer Science and Information Engineering, Tunghai University
Taichung, Taiwan, China
jwo@mail.thu.edu.tw

ABSTRACT

This paper introduces how to apply OMG's Model-Driven Architecture (MDA) to develop a cross-platform workflow solution. A workflow solution first is modeled as a PIM (Platform Independent Model) model. The PIM model will be translated into a PSM (Platform Specific Model) model according to the selected workflow platform and then the corresponding process definition is automatically generated. In this paper the result shows MDA can help preserve the knowledge of an application as a PIM model. The quality of the application also can be dramatically increased since the translations from PIM model to PSM model and from PSM model to the final production codes are automatically performed.

Keywords: e-Business, MDA, meta-model, WfMC, workflow management system

1. INTRODUCTION

In an enterprise workflow solution is one of the major building blocks for e-Business. In fact, workflow related solutions are widely studied recently [3, 4, 5, 7, 17, 19]. Among these researches, WfMC (Workflow Management Coalition) is the key organization to define the standard for workflow management system [4]. However, even with the standard, the migration of a workflow application from one platform to the other is still quite difficult since different workflow platforms have different philosophy to model their workflow processes.

MDA (Model-Driven Architecture) is a new technology for developing software applications proposed by OMG (Object Management Group) [12-15]. In the past, an application is modeled by also considering the IT platform and therefore the software design can not be reused when the platform is changed [1, 2, 8, 9, 18]. In MDA technology, an application is modeled only from the application logic without considering the IT platform [16]. This model is called a PIM (Platform Independent Model) model. When the IT platform for the application is decided, a corresponding model translated from the PIM model is created automatically. The new model is called a PSM (Platform Specific Model) model. Based on the PSM model and the translation rules, the production codes of the application are also generated automatically [10, 11]. Since the knowledge of the application is captured in the PIM model, it can be reused much easier later when the application needed to be migrated to another platform. The quality of the application is also dramatically increased since the translations from PIM model to PSM model and from PSM model to the final production codes are automatically performed.

In order to show the superiority of adopting MDA technology for building e-Business related solutions, we

propose a new workflow application development methodology based on MDA. Figure 1.1 is the illustration of the idea. We define a workflow PIM meta-model named MyWF meta-model. In this paper, we chose WfMC's XPDL process model as the experimental PSM meta-model. We also define the translation rules for MyWF model to the XPDL model and for XPDL model to the final XPDL process definition.

The organization of this paper is as follows. In section 2 we briefly introduce WfMC's XPDL model. MyWF meta-model is defined in Section 3. The implementation of the translation rules is described in Section 4. Section 5 is the conclusion remark of this paper.

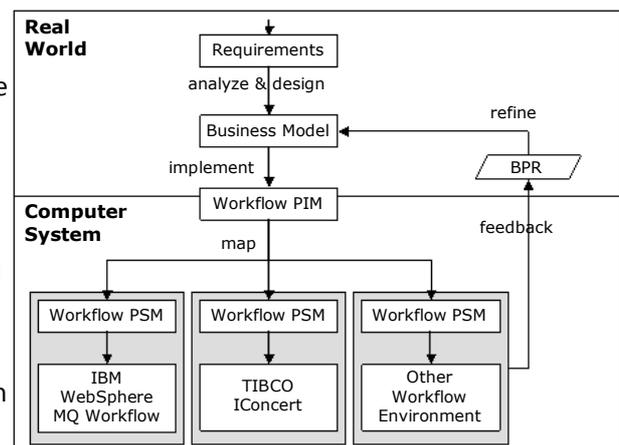


Figure 1.1. The architecture of applying MDA to build workflow application

2. XPDL

According to the workflow reference model proposed by WfMC, a workflow application has three different modes, namely, build-time functions, run-time control functions, and run-time interaction functions. Therefore, a workflow application can be considered as an

application with multiple operation steps such that it is routed automatically by following its corresponding process definition. The process definition standard of WfMC is called XPDL (XML Process Definition Language). XPDL is based on W3C's XML Schema [19]. Figures 2.1 and 2.2 are the meta-model and the package meta-model for XPDL, respectively.

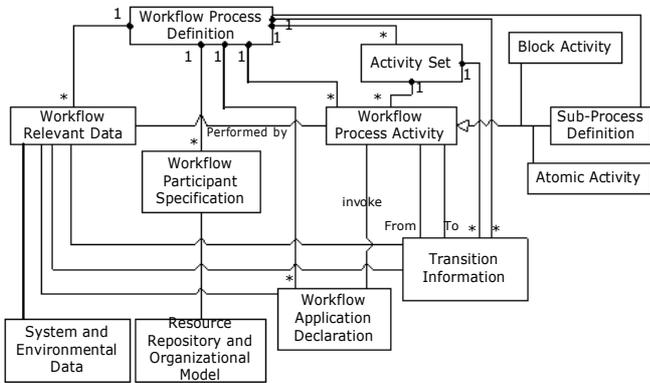


Figure 2.1. XPDL meta-model diagram.

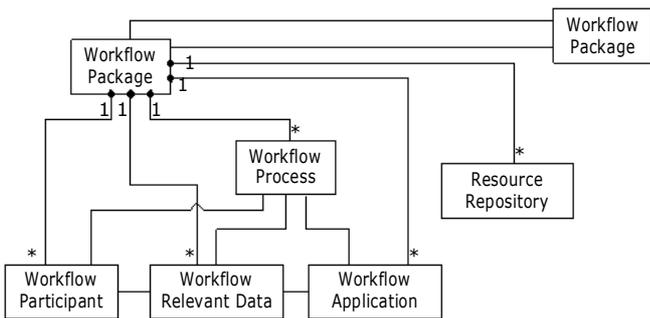


Figure 2.2. XPDL package meta-model diagram.

3. MyWF META-MODEL

In this section we define a generic workflow platform independent meta-model named MyWF. Figures 3.1 and 3.2 are the design of the MyWF meta-model. In this PIM model, there are five elements, namely, Process, Task, Transition, Resource and RelevantData. Process element is used to describe a workflow's operation steps. Task element is to store the information of the operation in each step. The information about the routing between two steps is defined in Transition element. The information of the resource required for the workflow application is stored in Resource element. RelevantData element contains the information across different steps. Each Process element can have more than one Task and various Transition, Resource and RelevantData elements. Task is subdivided into two different types, namely, GroupTasks and SubFlow elements. Resource could be either Application resource or user Role elements.

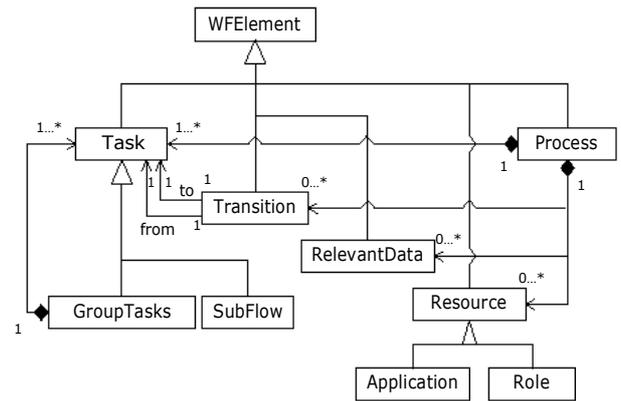


Figure 3.1. MyWF meta-model diagram.

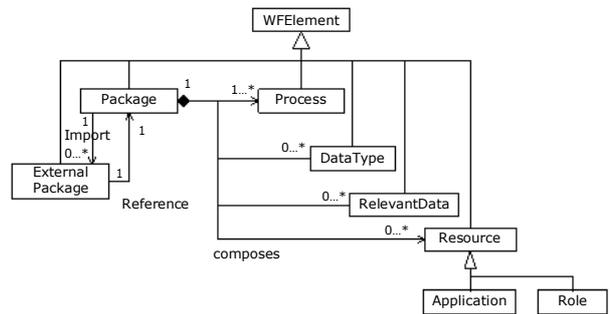


Figure 3.2. MyWF package meta-model diagram.

4. MyWF TO XPDL TO PRODUCTION CODE TRANSLATION

In the previous two sections, we have a generic workflow PIM model called MyWF meta-model and a well known workflow PSM model named XPDL meta-model. In this section, we describe how to transfer a MyWF application model into XPDL model and then transfer the XPDL model into its corresponding production codes.

4.1 MyWF PIM Model to XPDL PSM Model

The translation design between MyWF PIM model and XPDL PSM model is separated into two solutions. The first part is to guarantee that the traversal of the PIM model elements is complete. The second part is to define the translation rules among the elements between two models. We use Bean Script Framework (BSF) to implement the two solutions. The reasons of selecting BSF to build the MDA translation solution are:

- ò *Flexibility*: the interpreting based script language is more flexible when translation rules are changed and are required to be re-implemented.
- ò *Friendly implementation*: BSF is easier for MDA developers to use when implementing the translation rules among different models.

The translation algorithm is described as follows.

- (1). Retrieve the next element in the MyWF PIM model.
- (2). Identify the element type of the current element. If the type is unknown then throw `UnknownTypeException` and interrupt the traversal.
- (3). According to the translation rules, fetch and execute the corresponding script code and perform the translation. If the script is not found, throw `NoScriptException` and interrupt the translation step.
- (4). Retrieve the related translation required information and put it into the `ScriptEngine`.
- (5). Execute the corresponding script code to build the translated XPDL elements. If the execution can not be executed correctly, throw `TransformationException` and interrupt the execution.
- (6). Inspect the current MyWF element and make sure if there is any related element. If yes, go to step 1, otherwise go to step 7.
- (7). Return the translated XPDL elements. Go to step 1.

4.2 WfMC XPDL to Production Codes

In this section, we describe the translation rules between PSM model and the final production codes.

For each element in the PSM meta-model, we define one or more than one templates. Each template is a code pattern. The translation algorithm is as follows.

- (1). Retrieve the next element in the PSM model.
- (2). Identify the element type of the current element. If it is not known, throw `UnknownTypeException` and interrupt the execution.
- (3). Retrieve the corresponding templates. If there is no template is available, throw `NoTemplateException` and interrupt the execution.
- (4). Parsing the template pattern code. Identify the required information for the template from the element. If there is no corresponding information, throw `NoTemplateAttributesException` and interrupt the execution.
- (5). Return the generated production codes. Go to step

5. CONCLUSION

Figures 5.1 is the diagram to show how to define the roles and development process when applying MDA technology to build a workflow application. Figure 5.2 is the screenshot of our example implementation using Java and JMI [6]. From our experience of this research, we find that MDA indeed can greatly reduce the effort when implementing e-Business solutions on different platforms.

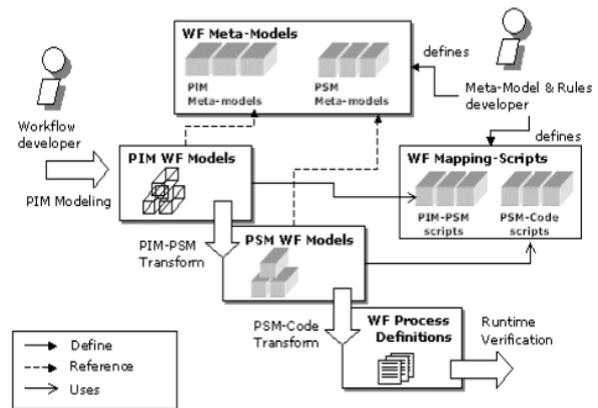


Figure 5.1. The roles and development process when applying MDA to workflow application.

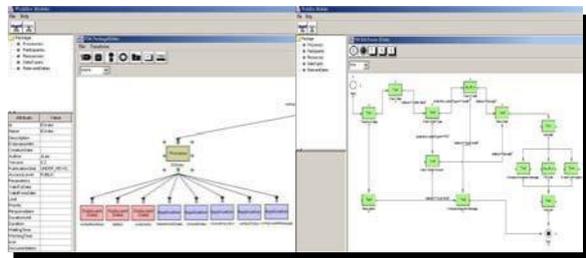


Figure 5.2. A screenshot of the example solution.

REFERENCES

- [1] Ahmed Abd-Allah, "Composing Heterogeneous Software Architectures," Doctoral Dissertation, Center for Software Engineering, University of Southern California, Los Angeles, CA 90089, August 1996.
- [2] Alessandro Bianchi, Danilo Caivano, Giuseppe Visaggio, "Iterative Reengineering of Legacy Systems," *IEEE Transactions on Software Engineering*, volume 29, issue 3, March 2003, pp. 225-241
- [3] Albano, F.; Pino, J.A.; Borges, M.R.S., "Participatory business process reengineering design: generating solutions," *IEEE SCCC 2001*, pp. 13 -22
- [4] David Hollingsworth, "Workflow Management Coalition The Workflow Reference Model," Document Number WfMC-TC-1003, 19-Jan-95, 1.1, <http://www.WfMC.org>
- [5] Diimitrios Georgakopoulos, Mark Hornick, Amit Sheth, "An Overview of Workflow Management: Form Process Modeling to Workflow Automation Infrastructure," *Distributed and Paralled Databases*, pp. 119-153
- [6] JMI Expert Group, "Java Metadata Interface (JMI) Specification," Version 1.0, 07-June-2002, <http://java.sun.com/products/jmi/>
- [7] Keith D. Swenson, Kent Irwin, "Workflow technology: trade-offs for business process re-engineering," *Conference on Supporting Group Work*, 1995, pp. 22-29
- [8] L. Davis, R. Gamble, and J. Payton, "The Impact of Component Architectures on Interoperability," *Journal*

of Systems and Software, Volume 61, Issue 1, March 2002, pp. 31-45.

[9] Ling Feng, Hongjun Lu, Allan Wong, "Integrating Legacy Systems Towards Intelligent Enterprise Computing," IEEE SMC '99 Conference Proceedings, Volume 2, 1999, pp. 184-189.

[10] Milicev, D.; "Automatic Model Transformations Using Extended UML Object Diagrams in Modeling Environment," IEEE Transactions on Software Engineering, volume 28, issue 4, April 2002, pp. 413-431.

[11] Oldevik, J.; Solberg, A.; Elvesæter, B.; Berre, A.J.; "Framework for model transformation and code generation," IEEE Enterprise Distributed Object Computing Conference, 2002, pp. 181-189.

[12] OMG, "Model Driven Architecture (MDA)," OMG Document number ormsc/2001-07-01, <http://www.omg.org>

[13] OMG, "Meta Object Facility Specification," OMG Document number formal/02-04-03, <http://www.omg.org>

[14] OMG, "XML Metadata Interchange Specification," OMG Document number formal/2002-01-01, <http://www.omg.org>

[15] OMG, "Unified Modeling Language Specification," OMG Document number formal/2003-03-01, <http://www.omg.org>

[16] Snoeck, M.; Dedene, G.; "Experiences with Object Oriented Model-Driven Development," IEEE Software Technology and Engineering Practice Proceedings, 1997, pp. 143-153.

[17] Schmidt M.T. , "The evolution of workflow standards," IEEE Concurrency, Volume: 7 Issue: 3, July-Sept. 1999, pp. 44-52

[18] Sakib Abdul Mondal, Kingshuk Das Gupta, "Choosing a middleware for web-integration of a legacy application," ACM SIGSOFT Software Engineering Notes, Volume 25, Issue 3, May 2000, pp. 50-53

[19] Workflow Management Coalition, "XML Process Definition Language Specification (Draft 1.0)," Document Number WfMC-TC-1025, <http://www.WfMC.org>

E-Alliances: Search and Match for Small Business Partnership

Alan Chonghua Guo

E-Alliances Research Center
Gyrox Business Research Inc., Beijing 100021, China
Tel: 86-10-51666689, Fax: 86-10-87783730
guoch@gyrox.com

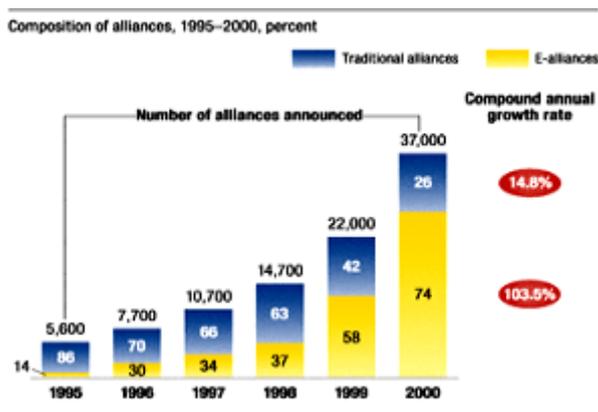
ABSTRACT

Companies across all sectors with increasingly smaller size rely on alliance relationships to bolster revenues during times of economic uncertainty. When forming alliances, large firms typically hire vendors such as investment banks to help identify potential alliance partners. However, small businesses with intention to build alliance find difficulties to search desired partners especially on a global arena due to three causes: 1) lack of skills, resources and information; 2) too small to capture the attention of investment banks or alike; 3) not affordable to the high search fee charged by vendors. Thus, reliable information about small business market is often limited. The odds of successful search and match become very low given their limited access to right information. In fact, traditionally labor-intensive search and match either by investment banks or by firms themselves have been both costly and time-consuming. On the demand side, small businesses are searching for more alliance opportunities than ever before while deal size has been increasingly declining since the last two decades. The globalizing, deregulation and internet are attributed to the driving forces of this trend. Under this circumstances, E-Alliances is able to close the huge gap between the demand side which represents the tremendous alliance search needs of small businesses worldwide and the supply side which is constrained by search capacity. In this paper, the author will interpret how E-Alliances is able to help worldwide small business companies to reach their desired partners in a fast, cheap and DIY manner. The author will also discuss the outlook of alliance.

Keywords: E-Alliances, Deal Fit Index, Aladdin Setting, Liquidity

1. INTRODUCTION

Companies across all sectors with increasingly smaller size rely on alliance relationships to bolster revenues during times of economic uncertainty. Indeed, the percent of total business involving alliances continues to grow rapidly over time. Alliances enable companies to pursue new growth opportunities without incurring the costs and risks associated with other development options.



Source: McKinsey 2001

Figure 1. Growth of alliance activities

When forming alliances, large firms typically hire vendors such as investment banks to help identify potential alliance partners. However, small businesses with intention to build alliance find difficulties to search

desired partners especially on a global arena due to three causes: 1) lack of skills, resources and information; 2) too small to capture the attention of investment banks or alike; 3) not affordable to the high search fee charged by vendors. Thus, reliable information about small business market is often limited. As a result, many companies in the small market have developed their own methods to identify and obtain information concerning potential alliance targets, rarely engaging investment banks during the identification phase, but occasionally employing them as the deal progresses. The odds of successful search and match become very low given their limited access to right information. In fact, traditionally labor-intensive search and match either by investment banks or by firms themselves have been both costly and time-consuming. To a large extent, this explains why investment banks are lured by large deals while rejecting small ones. On the demand side, small businesses are searching for more alliance opportunities than ever before while deal size has been increasingly declining since the last two decades. The globalizing, deregulation and internet are attributed to the driving forces of this trend. Under this circumstances, E-Alliances is able to close the huge gap between the demand side which represents the tremendous alliance search needs of small businesses worldwide and the supply side which is constrained by search capacity. The purpose of this paper is to examine how E-Alliances is able to help worldwide small business companies to reach their desired partners in a fast, cheap and DIY manner. The author will also

discuss the outlook of alliance. The results suggest that the entry of e-Alliances should stimulate alliance activity and enhance the efficiency of market.

In an alliance market, there exist huge numbers of firms which look for alliance opportunities. E-Alliances provides an online marketplace for firms to search and match each others. Each firm will first register in by offering its company information and desired partner criteria through a series of simple click of either single or multiple choices. To obtain an even higher quality partner list, searching firms can even enter more selection criteria by category method such as alliance type, locations, industry and/or by weighted preference of criteria called Deal Fit Index (DFI). The criteria designated basically enables specific firm to sort potential partners according to its alliance objectives and appropriate alliance structures which varies from equity-based alliances such as merger & acquisition, joint venture, etc. to non-equity-based ones such as joint-R&D, license agreement, OEM, etc. With all individual firm's information inside, E-Alliances automatically matches the firms and provides a shortlist of potential partners for each member firm. For advanced members, they are able to view a highly selective shortlist based on DFI, which are sorted from a huge member database throughout the world. The search and match of E-Alliances are 7×24 real-time based which automatically send email to logged-out members to inform them of new prospects which show interest in them. E-Alliances also provides mailbox for members to connect each other for further relationship building. The author concludes that E-Alliances significantly improves firms' ability to search quality partners from both scale and scope in a real time, global reach, and cost effective way. To facilitate the deal progress and provide the members with total solution, E-Alliances will strive to create the value chain by providing customer education programs to develop members' alliance skills and inviting qualified vendors to provide professional services like due diligence, valuation, negotiation, accounting and law, post-alliance integration, etc.

2. PERFORMANCE OF STRATEGIC ALLIANCE

Research has reported that building alliance will improve performance of both allying firms and market with an example of OEM partnership^[1]. (Ren 2003) Consider a simple economy with only two firms: firm A and firm B. The both firms make same product. Firm A is legally allowed to sell its product in region B; in symmetry, firm B is legally allowed to sell its product in region A. Firm A and firm B can only make decision on the quantity of their own production, i.e., Cartel is prohibited.

The product price is the function of production in both region A and B, denoted by the following equation (1):

$$P(z) = a - z \tag{1}$$

The marginal cost is denoted by c and suppose there is no fixed cost of production. However, a firm must incur a fixed a fixed cost of distribution channel management in order to sell in a region, denoted by D.

Furthermore, in order to sell in an outside region, a firm has to not only incur the distribution cost but also bear the shipment cost. This suggests that the total cost of making and shipping one unit product is c/g with 1 < g ≤ 1 (shipment cost is zero if g=1). Q is the rental cost of one firm pays to another for the use of its infrastructure and distribution channel.

In direct competition, the two firms pursue each own profit maximization as equation (2) and (3):

Firm A:
$$Max_{x,x} (x p_{(x+y)} + x^* p_{(x^*,y^*)} - cx - \frac{c}{g} x^* - 2D) \tag{2}$$

Firm B:
$$Max_{y,y} (y p_{(x+y)} + y^* p_{(x^*,y^*)} - cy - \frac{c}{g} y^* - 2D) \tag{3}$$

On the other hand, when two firms ally in OEM model, they agree on the equilibrium point for a rental value q. Because the two markets are symmetric, we only consider the region A. In market region A, the profit maximization conditions of two firms are as shown below:

Firm A:
$$Max_x (x p_{(x+y)} - c) - D + y(q - c) \tag{4}$$

Firm B:
$$Max_y (y p_{(x+y)} - q) \tag{5}$$

To evaluate market performance, Ren (2003) examined the respective impact of the above two models on total surplus of welfare. In the Nash equilibrium of direct competition in market region A, the total surplus of welfare is as below:

$$W_1 = [\int_{z_{e1}}^a p_z dz - z_{e1} p_{z_{e1}}] + [z_{e1} p_{z_{e1}} - c x_{e1} - (c/g) y_{e1} - 2D] \tag{6}$$

In the Nash equilibrium of OEM alliance model in market region A, the total surplus of welfare is as below:

$$W_1 = [\int_{z_{e2}}^a p_z dz - z_{e2} p_{z_{e2}}] + [z_{e2} p_{z_{e2}} - c x_{e2} - c z_{e2} - D] \tag{7}$$

Obviously, W₁ and W₂ are the functions of the fixed cost D to set up distribution channels. To compare the performance of the above two situations, we find the partial derivative of the difference between W₁ and W₂ against variable D. as shown below:

$$\frac{\partial(W_1 - W_2)}{\partial D} = \frac{\partial(W_1 - W_2)}{\partial q} \cdot \frac{\partial q}{\partial D} = \frac{c - a + 9\sqrt{D_{max} - D}}{8\sqrt{D_{max} - D}} \tag{8}$$

Ren (2003) further discussed how W₁ - W₂ differs in two cases when D=D_{max} and D=0, respectively. He drew the conclusion that OEM alliance helps increase the total welfare and thus improve the market efficiency.

Studies of alliance performance are discussed in many other business literatures. Jay^[2] (2002) suggested that alliance relationship building can be considered as real option so that its performance could be valued through adjusted Black-Scholes equation. However, he did not illustrate how to do that but instead, went back to a more intuitive NPV rule to value the synergy of allying firm A and firm B as below:

$$NPV(A+B) > NPV(A) + NPV(B) \quad (9)$$

3. ALLIANCE SEARCH AND MATCH THROUGH INTERMEDIARIES

3.1 Cost of direct search and intermediary search

Consider a market with many heterogeneous firms who search and match for alliance. We review a simple economy with two firms: Firm A and Firm B, to seek each other for partnership. Before alliance, the two firms maximize their profit V_A and V_B , respectively. They directly reach each other to bargain and finally decide to ally. After allying, they achieve a profit V . However, a search cost is C . We suppose the alliance brings in a positive gain for both parties as below:

$$V - C > V(A) + V(B) \quad (10)$$

For simplicity, we suppose the both parties equally share the gain:

$$\frac{V - C}{2} \quad (11)$$

Given the uncertainty for search directly, suppose an intermediary can help both firms to find each other. The intermediary incur the search cost T . Apparently, when and only when the intermediary can reduce the search cost, the indirect transaction through intermediary can really happen, that is

$$T \leq C \quad (12)$$

The deal volume processed by intermediary is far exceeding that of an individual firm. Moreover, an intermediary stay much longer in the market than any single firm and thus build up a credibility which lasts much longer than that of an individual firm. The deal volume and brand longevity of intermediary create cost advantage over the direct search.

3.2 Cost of human search and internet search

The search by intermediary can be human-based historically (e.g., investment banks, accounting firms, etc) and web-based which is innovative E-Alliances. Because of the uncertainty of search business, both human search and internet search will incur search cost. For instance, search results via internet can immediately appear on the screen while human search must wait for longer time. After wait, the alliance gain which is $V - T$ is achieved and equally shared between the allying two firms. If the two firms discount the future alliance gain with ratio δ , then the present value of the human search equals to

$$\frac{\delta(V - T)}{2} \quad (13)$$

This implies that the human search cost is

$$H = (1 - \delta)(V - T) \quad (14)$$

When and only when

$$I \leq H = (1 - \delta)(V - T) \quad (15)$$

The internet search is better off than the human search. Since the human reach is limited by time and space, the human search can find the right match at a probability β . Suppose the website of internet search is widely known, thus the human search cost will be

$$H = (1 - \beta)(V - T) \quad (16)$$

If the matching process by human search is inefficient, then β is very small; consequently, internet search will be superior to the human search. In reality, while big firms have a broad network, professional expertise and good money to guarantee efficient search (increase β), small business is inferior to this situation with limited access, few talents and less money to lure attention of search vendors (β is fairly low). Accordingly, a new approach to search and match for small business partnership is essential. E-Alliances becomes appealing to alliance search for small business due to its self-service, global reach and low cost.

4. WORKING MODEL OF E-ALLIANCES

E-Alliances provides an internet search marketplace for firms to search and match ideal partners. In particular, small business firms are targeted customers for this marketplace with driving factors in three folds: 1) neglect by human search institution because the deal is too small to capture the attention of investment banks or alike; 2) inaffordability to high human search cost; 3) liquidity generated by positive feedback of internet which follows the Metcalf Law. The online alliance search engine is supposed to provide small business a perfect tool to search and match at low cost, high speed and global reach.

4.1 Quick match through Aladdin setting

To increase liquidity by attracting as many as possible firms to join in paid-membership, E-Alliances develops a quick match functionality to quickly show real-time match results by asking search firms to create their own Aladdin setting, which requests the firm to make single or multiple choices against a designated questions reflecting a series of search criteria of the firm. Once the firm finish and submit its result, it instantly shows on the screen a list of prospects which satisfy its Aladdin settings, as shown below



Figure 2. Illustrated list of quick match

The sequence of prospect list is sorted in Basic Match status based on their match level of criteria preset in Aladdin setting. The search firm can further review a prospect’s profile by clicking its username as shown below



Figure 3. Illustrated prospect profile

If the firm is interested in the prospect, it can either send a email to the prospect or add the prospect into its hot list for later contact.

4.2 Browse profiles through demographics

To satisfy firms’ need to search by demographics, E-Alliances provides category search such as alliance type, geographic and industry. The prospect list then further sequenced according to matching level with the above demographics, as shown below



Under this search mode, search is more accurate and time is further saved for the member firms.

4.3 Advanced search through Deal Fit Index®

To identify potential partners, searching firms should formulate the criteria for evaluating potential partners, and determine which candidate best meets these criteria. The main goal of the firm is to determine if there are suitable partners that can help it fill its capability gap. Making this selection requires that the firm understand and rank each prospective partner's critical attributes in terms of how well they meet the firm's criteria. E-Alliances provides the paid members with a designated criteria with weighted preference through single or multiple choices by simple clicks. As result of this exercise, the firm can quickly present the conceptual structure of the alliance to the candidates for their matching criteria. At the end of this stage, the firm should have a shortlist of prospects with best fit on its objectives for entering into the alliance, which are ranked by their Deal Fit Index (DFI®). The firm may also wish to re-examine the list of required partner attributes to confirm that the criteria is realistic by going back to its original setting to make adjustment of weight and preference.

5. CHALLENGES FOR E-ALLIANCES

Besides the functionalities designated, three other factors appear intrinsic to any marketplace able to sustain itself over the long term: liquidity, credibility and value-added services. These elements determine whether a marketplace can enlist users and provide them sufficient value to continue their participation. A marketplace initiative must manage these key issues well.

5.1 Liquidity

Generating liquidity by attracting enough customers is the most significant challenge to a marketplace’s success. A new, electronic marketplace must compete against traditional methods of alliance service. While the e-business may offer administrative savings through automation, the presence of too few participants will leave marketplace prices out of alignment with regular channels.

5.2 Credibility

As a new, shared model for conducting business, marketplaces depend upon the trust of their participants. The marketplace host maintains access, at least theoretically, to all information that parties exchange. This insight into pricing and purchasing behavior can scare off participants fearful that the information may be turned against them in some way. While regulators will set rules limiting the worst abuses, consortia must take additional steps to demonstrate neutrality and build trust among marketplace users. Methods include setting up independent management executives and information firewalls between owners and participants. Pure plays

can attempt to claim the high ground as disinterested parties because the owners never buy or sell goods in the marketplace. However, as pure plays begin to accept equity investments from industry incumbents, the lines between ownership models begin to blur. Yet, in the end, liquidity cannot develop if users mistrust the marketplace and conduct transactions on other sites or through other channels.

5.3 Value-added services

To succeed in a long run, E-alliances must effectively integrate the value chain services into an alliance solution to the client.

Regardless of the alliance structure, alliance process is inherently complex at variety of difficulties which are far beyond small firms' capability because it heavily involve in people, process and tools. In other words, E-Alliances alone is not sufficient to satisfy firms' alliance needs. Alliance services typically comprise the following phases:

Alliance needs assessment. This services encompasses strategy consulting, including needs assessment, assistance with a decision on whether alliance is a feasible and beneficial option, evaluation of various options (m&a, joint venture, divestiture, OEM, channel sales agreement, joint R&D, etc.), and assistance in choosing the right option. The strategy work requires a deep knowledge of the client's business and tight alignment of merger options with the overall short- and long-term business strategy of a company.

Partner selection. The imperative task in this phase is assessing the strategic value of the partner. It is important not to treat it as a standalone activity. The valuation process must be closely integrated with the business strategy and integration plan to achieve the desired result.

Due diligence. Due diligence is conducted by the lead alliance services provider, or it is outsourced to an independent party.

Transactional services. Transactional services include negotiation and financial/legal advisory. A partner is chosen either by the alliance services provider together with the client.

Customer education services. Customer education programs seek to deliver information that enables customers to more effectively utilize a company's product.

6. OUTLOOK OF E-ALLIANCES

So far, I have interpreted how E-Alliances will help small business address their critical alliance issues from a market microstructure perspective. Since E-Alliances has intrinsically a global orientation, both from an alliance point view and internet standpoint, we are optimistic that it will become the world's largest base of

alliance. Its success will be not only measured by hit rate of customers, but also by the alliance deal success rate. Furthermore, E-Alliances provides an emerging market like China with a unique value to attract foreign direct investment, facilitate knowledge and technology transfer, and boost trade volume among countries. It will also help stimulate information transparency and establishment of national credit system. In essence, E-Alliances will contribute to accelerate the growth of worldwide small business by fostering successful alliances among them wherever they are.

7. CONCLUSIONS

Firms with complementary value are more likely to achieve synergetic performance through alliance than that of they work alone, even though the success is not guaranteed (according to David and Tammy^[3] 2000, 50% of the alliance failed). To seek a partner, a firm can search by itself or by use of intermediary. The decision depends on the cost incurred by each alternative. For intermediary search, internet search reshaping the alliance economics by competing with the traditional human search. Giving its self-service, low cost, speed, global reach and 7×24 working mode, internet is gaining competitive advantage over the human search. Small business firms can benefit from the internet search most for they have limited access, few talents and less money. As an online marketplace, E-Alliances provides the small business with variety of search approaches to meet their search needs for alliances. However, while E-Alliances offers creative, promising functionalities, it faces challenges of liquidity and credibility generation, too. In particular, E-Alliances' success largely relies on the ability to restructure the alliance value chain by integrating E-Alliances and necessary value-added services into customer solution. To look ahead, E-Alliances will become the world's largest base of alliance and contribute to both micro- and macro economy by facilitating firm collaboration around the world.

REFERENCES

- [1] Jianxin Ren, *Research of corporate strategic alliances*, China Finance & Economics Publishing, Beijing, 2003.
- [2] Barney, J.B., *Gaining and sustaining competitive advantage*, Pearson Education Asia Limited, 2002
- [3] Ernst D., Halevy T., "When to think alliance", *The McKinsey Quarterly*, Vol. 1, pp77-83, Beijing, 2001.
- [4] Spulber, D.F., *Market Microstructure: intermediaries and the theory of the firm*, Cambridge University Press, 1999

Exploring Sources of Competitive Advantages in E-business Application in Mainland Chinese Real Estate Industry

Yan Tao¹, C. Matt Hinton²

¹The Open University Business School, Walton Hall, Milton Keynes, MK7 6AA, UK,
y.tao@open.ac.uk, C.M.Hinton@open.ac.uk

ABSTRACT

One of the key issues in e-business research is how established companies can gain competitive advantage by exploring e-business. Despite the interest in e-business applications by traditional firms, few empirical studies have been carried out to look at how ‘clicks-and-mortar’ approaches offer competitive advantages, especially from a specific industry perspective. This study investigates the key sources of competitive advantage gained from e-business applications by Chinese real estate developers and whether the value chain theory and its related theories can explain this phenomenon. By using a qualitative case study approach, the study shows that the value chain framework is useful to identify and categorize possible e-business application areas. Moreover, this categorization makes identification of key sources of competitive advantage explicit. However, this framework cannot fully explain the success of e-business applications nor the realization of intended motivations. Hence, further research and work are needed to make the value chain model become an easily-used, practical guideline for e-business implementation.

Key words: e-business, competitive advantage, value chain, real estate industry

1. INTRODUCTION

This study investigates the key sources of competitive advantages gained from e-business applications made by Chinese real estate developers and whether the value chain theory and its related theories can explain this phenomenon.

One of the key issues in e-business research is how established companies can gain competitive advantage by exploring e-business. Despite the interest to e-business application by traditional firms, few empirical studies have been carried out to look at how ‘clicks-and-mortar’ approaches offer competitive advantages, especially from specific industry perspective (Steinfeld *et al.*, 2002).

Therefore, the aim of this research is to address this gap in the current literature by investigating sources of competitive advantage gained from e-business adoption by established real estate developers in Mainland China.

This aim can be broken down into the following questions:

- 1) What are the key sources of competitive advantage gained from e-business application by real estate developers?
 - 2) How effectively can the value chain model capture the key sources of competitive advantage?
 - 3) Does the transformation to value chain made by some researchers pass the empirical tests in specific industry?
- Or,
- 4) Can any new framework/theory of transformation arise from empirical data analysis?
 - 5) Do any links exist between value chain framework explanation to sources of competitive advantage in e-business and other alternative explanations?

2. RESEARCH METHOD

A case study and qualitative research approach will be adopted in this research. Following the suggestions given by Yin (1994), this research design will be based on multiple cases and multiple evidences.

Large spatial diversity between Chinese real estate developers, and time and resource limitations would suggest a strategy initially focused on established real estate development firms whose main activities focus on three geographical locations, namely Beijing, Shanghai, and Guangdong Province, because these cities are the most prosperous housing markets in Mainland China.

According to the e-commerce definition, that is “e-commerce is the sharing of business information, maintaining business relationships, and conducting business transactions by means of tele-communication networks” (Zwass,1996), adopted in this research, we define established real estate developers with e-business application as the company who conducts one or more than one of the following activities:

- the buy-side e-business transactions with suppliers;
 - the sell-side e-business transactions with customers;
 - the use of Internet technologies within the firm.
- Here, the key business activities refer to the nine generic categories of interdependent activities which compose a firm’s value chain. For example, Floris *et al.* (2001) give an insight into the reach of e-business technologies into the value activities. (Shown in **Figure 1**)

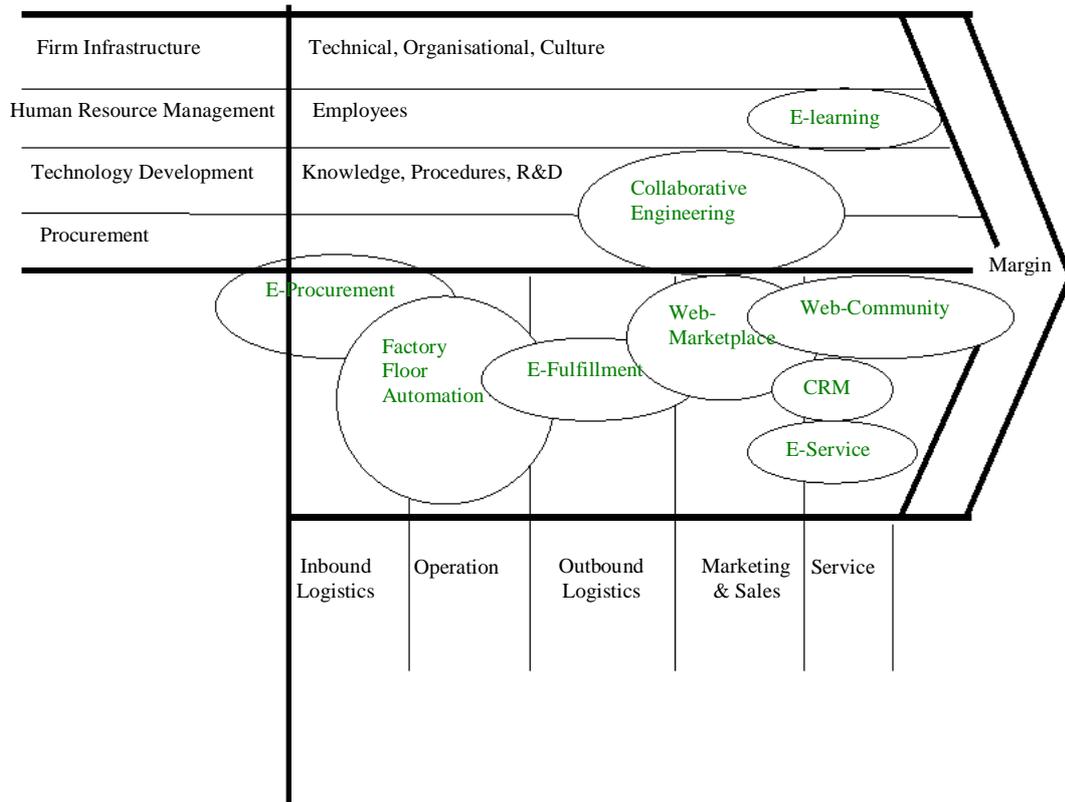


Figure 1 E-business technologies applications in the value chain

Source: Floris *et al.* (2001)

Cases were identified by combining the databases of real estate developers with first-grade qualification, top ten real estate developers in 2002 and top 35 public real estate development companies in 2001. Additional criteria were used which identified companies with continuously better than average performance, an e-business application, and geographic focus. Twenty companies were initially chosen and their websites scanned to confirm whether they are using e-business. The initial finding is that most companies in the database only use websites as e-brochures and only a few companies have used e-business actively. This narrows down the cases into five companies, who are located in Beijing or Shenzhen.

The data analysis is influenced by grounded analysis, four principles of high quality analysis given by Yin (1994): namely use all relevant evidence, consider all

major rival interpretations, address the most significant aspects of the case and use your own prior expert knowledge and value chain theoretical framework.

3. RESULTS AND ANALYSIS

The five case companies are Mainland China-based real estate development companies, among them three in Beijing and two in Shenzhen. All of the companies are well established in the real estate industry with property development as their core businesses and have, to some extent, tapped e-business. Some key features of the case companies and their use of e-business are summarised in **Table 1**.

A cross-case comparison on the key features from each case for is shown in **Table 2**.

Table 1 Main features of the case companies

Company	Market segmentation	Organizational structure	E-business application
I	Mainly middle-end; partly high-end Focuses on multiple cities with marketing expansion potential	Centralizes management cross regions	1) Marketing promotion and advertisement; 2) E-procurement; 3) Intranet for email system; 4) Automation of internal management processes.
II	Mainly middle-end; partly high-end Focuses on Shenzhen, Shanghai, and Beijing with marketing expansion potential	Centralizes management cross regions	1) Marketing promotion and advertisement; 2) Business to Employee platform for knowledge management; 3) Automation of internal management processes; 4) Customers' e-forum.
III	Mainly middle-end; partly high-end Focuses on Beijing with marketing expansion potential		1) Marketing promotion and advertisement; 2) Automation of internal management processes; 3) Secondary housing market.
IV	High-end Focuses on multiple cities with marketing expansion potential		1) Marketing promotion and advertisement; 2) B2C platform to allow customers order houses.
V	Mainly low and middle-end; partly high end; focuses on multiple cities	Independent management by subsidiary companies	1) Marketing promotion and advertisement; 2) Intranet and financial system.

Table 2 Cross-case comparison

Case	Definition to e-business	E-business adoption motivations	E-business application	Advantages gained from application	Main concerns of E-business adoption
I	No clear idea but thinks it has no big difference with preceding information systems	1) To Share information; 2) To improve operation efficiency; 3) To reduce procurement cost; 4) Marketing promotion.	1) Marketing promotion and advertisement; 2) E-procurement; 3) Intranet for email system; 4) Automation of internal management processes.	1) Better information sharing; 2) Reducing purchasing cost; 3) Facilitating employee training; 4) Standardising operation process; 5) Reducing redundancy;	Integrate different information systems adopted by different functional departments.
II	E-business including Web-based transaction. Definition is not important to them.	1) To Share information; 2) To manage customer relationship and achieve strong interactions with customers; 3) Marketing promotion.	1) Marketing promotion and advertisement; 2) Business to Employee platform for knowledge management; 3) Automation of internal management processes; 4) Customers' e-forum.	1) Reducing redundancy; 2) Facilitating employee training; 3) Promoting innovation; 4) Better management through customer monitoring; 5) Better information sharing.	
III	E-business including Web-based transaction.	1) Marketing promotion; 2) To improve management efficiency;	1) Marketing promotion and advertisement; 2) Automation of internal management processes; 3) Secondary housing e-market.	1) Better information sharing; 2) Facilitating strategy formulation.	Integrate different information systems adopted by different functional departments.
IV		To facilitate the implementation of the company's management concepts	1) Marketing promotion and advertisement; 2) B2C platform to allow customers order houses.	1) Reducing business risk. 2) Strong interaction with customers.	
V	E-business is a new technology facilitated tool of transaction and providing transaction-related service	To redefine the value chain and standardizing real estate development process	1) Marketing promotion and advertisement; 2) Intranet and financial system.		Organization-al structure makes difficult to implement e-business

The following key issues emerge from the analysis of the cases.

1) Main e-business applications by Chinese real estate development companies;

- 2) Main motivations and benefits of e-business application;
- 3) The main hurdles of e-business application.

Main e-business applications by Chinese real estate development companies include e-procurement, automation of the internal management process and marketing efficiency.

The companies' motivations for e-business adoption initially arise from trying to solve key management problems, especially operational problems, by using e-business applications such as e-markets, intra- or interorganizational systems, and providing customer service. Another common motivation identified is sharing information between internal users (shown in **Table 2**). Additionally, the adoption of information systems is also motivated by cost saving. A further motivation is to achieve strong customer interactions.

Table 2 shows the main advantages gained from e-business application.

This advantages can be categorized according to Porter and Millar (1985)'s statement that information technology is permeating the value chain at every point by:

- 1) transforming the way value activities are performed;
- 2) transforming the nature of the linkages among value activities;
- 3) affecting competitive scope.

Figure 2 illustrates the main transforming of the way value activities are performed.

By adopting e-business to perform these activities, the case companies can achieve the competitive advantages listed in **Table 2**. **Table 3** illustrates the relationship between value activities and competitive advantages achieved.

The main problems of e-business application include internal interorganizational systems integration and the impact of organizational context such as organizational structure, management style and the customer base.

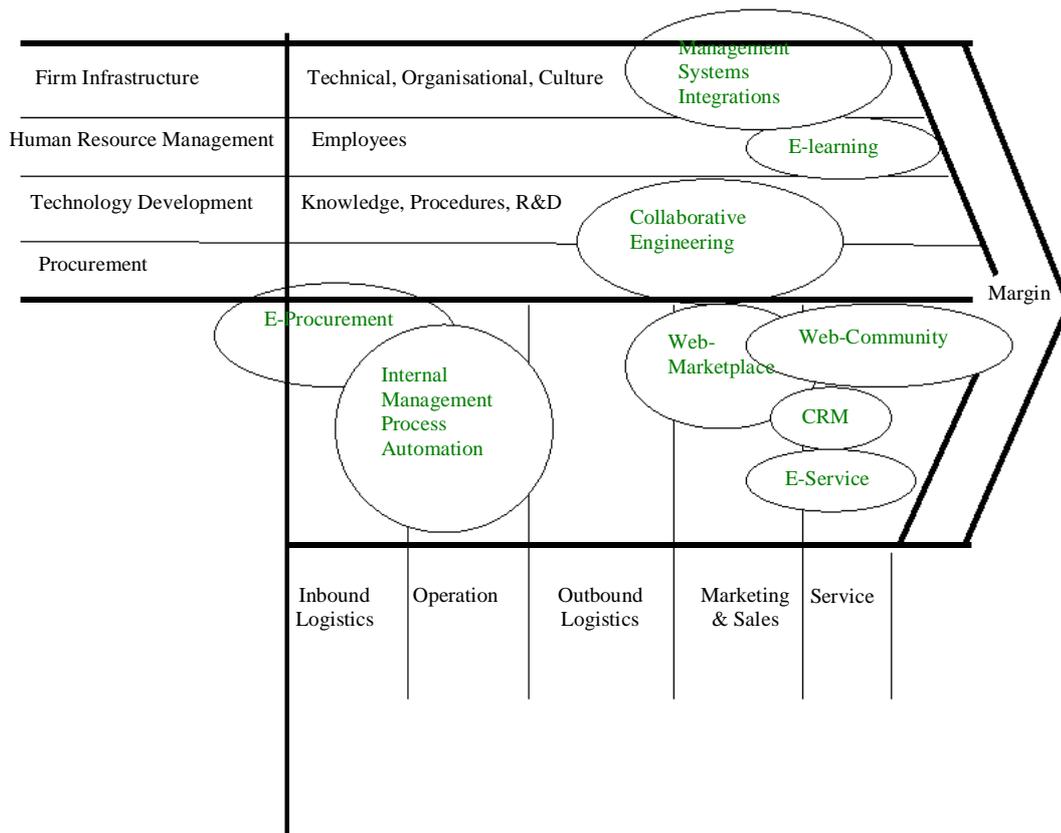


Figure 2 E-business technologies applications in the value chain in Chinese real estate industry
 Source: Adapted from Floris *et al.* (2001)

Table 3 Relationship between value activities change and competitive advantages achieved

Value activities	Adopted by companies	Competitive advantages achieved	Advantages achieved by companies
E-procurement	I	1) Reducing purchasing cost; 2) Standardizing operation process which leads to simplifying complex internal processes and reducing costs; 3) Reducing redundancy.	I
Internal management process automation	I,II,III,IV, V	1) Better information sharing; 2) Reducing redundancy.	I,II,III
Web-marketing	III (has intention to adopt, has not adopted yet)		
Service including e-service, CRM, Web-community	I,II, IV	1) Better management through customer monitoring; 2) Strong interaction with customers; 3) Reducing business risk.	II—better management; IV—strong customer interaction and reducing business risk
E-learning	I,II	1) Facilitating employee training; 2) Reducing redundancy.	I,II
Management systems integrations	III	Facilitating strategy formulation through better information sharing.	III

4. CONCLUSIONS

As **Figure 2** shows, in general, the value chain framework is useful to identify and categorize possible e-business application areas. Moreover, this categorization makes identification of key sources of competitive advantage explicit as **Table 3** has illustrated. However, this framework cannot fully explain the success of e-business applications and the realization of intended motivations for example the impacts of organizational structure, management style, and the customers base.

Floris *et al.* (2001)'s framework is instructive in asking interview questions about possible e-business applications and identifying possible e-business application areas. However, one common application, namely automation of the internal management processes, is not stated in this framework although it is similar to Factory Floor Automation.

Evans and Wurster (1999)'s suggestion of using e-business to resolve the trade-off between reach and richness helps in interpreting Company IV's e-business applications and competitive advantage gained from these applications.

Steinfeld *et al.* (2002) and Daniel and Grimshaw (2002)'s frameworks help in identifying and interpreting the motivations of e-business application and benefits achieved from e-business adoption.

Based on Floris *et al.* (2001)'s framework, **Figure 2** gives categorization of e-business applications adopted by established, big Chinese real estate development companies.

The findings illustrated in **Table 3** reinforce Porter (1985, 1998)'s finding that key sources of competitive advantages come from: redefining the role of traditional activities, optimising the linkages, and choosing competitive scope. This finding tends to support Porter (2001)'s claims that the fundamentals of economics and strategy have not changed and the underlying assumption that Internet is the latest stage in the ongoing evolution of information technology is valid.

However, further research and work are needed to make the value chain model become an easily-used practical guideline for e-business implementation. Practically, few case companies use the value chain model as theoretical instruction for their e-business experiment although Company I has used value chain theory to redefine its value chain but not for e-business application purposes. This is partly because of the reason described by Stabell and Fjeldstad (1998) that the value chain model is more suitable for the analysis of production and manufacturing firms than for service firms where the resulting chain does not fully capture the essence of the value chain. Another reason may be more explanation work needed to be done to make the value chain model more easily understood by enterprise managers.

Facilitated by government policy, as the Chinese real estate market becomes more and more transparent, it is anticipated that e-business will play a more and more important role in helping companies to achieve these three advantages. However, this conclusion is contrary to views given by most of the interviewees. The main reason for this is that they are working with different definitions of e-business. According to them, e-business must include Web-based transaction. However, in this

research, we would like to adopt the definition given by Zwass (1996): "E-commerce is the sharing of business information, maintaining business relationships, and conducting business transactions by means of telecommunication networks."

In summary, through case studies, the study shows that the value chain is helpful and primary in terms of understanding the different kinds of e-business applications in the practical world.

ACKNOWLEDGEMENT

The authors thank all the case companies for providing valuable information for this research.

REFERENCE

- [1] Daniel, E. M. and Grimshaw, D. J. (2002) 'An exploratory comparison of electronic commerce adoption in large and small enterprises', *Journal of Information Technology*, 17 (3), pp. 133-147.
- [2] Evans, P. and Wurster, T. S. (1999) 'Getting Real About Virtual Commerce', *Harvard Business Review*, 77 (6), pp. 84-94.
- [3] Floris, P. C., Hooft, V. and Stegwee, R. (2001) 'E-business strategy: how to benefit from a hype', *Logistics Information Management*, 14 (1/2), pp. 44-54.
- [4] Porter, M. E. (1985) *Competitive Advantage: Creating and Sustaining Superior Performance*, The Free Press, NY.
- [5] Porter, M. E. (1998) *Competitive Advantage: Creating and Sustaining Superior Performance-With a New Introduction*, Free Press, New York.
- [6] Porter, M. E. (2001) 'Strategy and the Internet', *Harvard Business Review*, March, pp. 63-78.
- [7] Porter, M. E. and Millar, V. E. (1985) 'How information gives you a competitive advantage', *Harvard Business Review*, 63 (4), pp. 149-160.
- [8] Stabell, C. B. and Fjeldstad, O. D. (1998) 'Configuring value for competitive advantage: On chains, shops, and networks', *Strategic Management Journal*, 19 (5), pp. 413-437.
- [9] Steinfield, C., Bouwman, H. and Adelaar, T. (2002) 'The dynamics of click-and mortar electronic commerce: opportunities and management strategies', *International Journal of Electronic Commerce*, 7 (1), pp. 93-119.
- [10] Yin, R. (1994) *Case Study Research: Design and Methods*, (2nd Edn), Sage, Thousand Oaks.
- [11] Zwass, V. (1996) 'Electronic commerce: structure and issues', *International Journal of Electronic Commerce*, 1 (1), pp. 3-23.

Research on Collaborative Commerce Model Based on Web Services

Liwen Wu, Yingqiang Duan

School of management, Dalian University of Technology, Dalian, Liaoning 116024, China
feagleme@sina.com

ABSTRACT

The concept of collaborative commerce was put forward by Gartner Group in 1999, and Web Services becomes one of the best implementation technologies of collaborative commerce for that it is loose-coupled, based on open standards and etc. In this paper, based on service oriented architecture of Web Services, we construct collaborative commerce model, analyze business process integration of collaborative commerce and design the implementation way of conversation mechanism in transaction process.

Keywords: Web Services, Collaborative Commerce, Transaction Process

1. INTRODUCTION

Collaborative Commerce is a new business mode on the background of global economics becoming one whole body. It can make enterprises easily collaborate with each other in the supply chain or cross supply chain. As result, the management mode is changed and resource is used economically^[1]. Collaborative Commerce is an evolution of the rudimentary point-to-point B2B transaction and it is a multi-point interaction among multiple parties for a single transaction.

Since the concept of collaborative commerce was put forward by Gartner Group, its implementation was researched from different aspects^[1]. Now, there are two partly implementation ways of collaborative commerce. One is web enabled c-commerce, which is a one dimensional, single e-enabled business process that allows certain internal data to be visible to external trading partners, and vice versa. Typically, this meant implementing a web presentation of the data, and allowing partners access to it. This is a limited form of c-commerce, with very limited value, saving only labor cost. Typical applications would be displaying demand for production materials, showing sales forecasts to suppliers, or presenting bills electronically. The other is B2B exchanges in markets such as steel, auto parts, chemicals or airline equipment. Buyers, sellers and suppliers are integrated through a web portal. In this stage, each enterprise can reconfigure its supply chain through the marketplace to leverage aggregated buying power and eliminate brokerage fees and middlemen^[2].

Those two ways of C-Commerce have never reached the critical mass required for mass adoption by all industries, because of the cost of integration. Information systems of two sides in the transaction are cannot communicate directly and isolated information islands are formed between them. And also, one complete transaction not only includes the two transaction sides, but also needs the third parties such as logistics, bank and etc. How to incorporate the third parties value-added service providers into business process and communicate each other between different systems freely is critical to

completely implement collaborative commerce, and one effective way is needed in order to implement transaction process integration.

Now, there are mainly two ways on implementing collaborative commerce. One is expanding the existed information systems such as SAP, the other is by adopting distributed component technology^[3]. But these methods have different defects. By expanding the system function, it can partly improve business process, but it cannot satisfy the need of easily process integration and reengineer, especially between heterogeneous systems. Distributed component technologies, such as DCOM, CORBA and RMI, are complex, their message format are not open, and are difficult through firewall. So, they cannot satisfy the demand of transaction process integration of collaborative commerce.

Web Services, which are loosely coupled, cross-internet, based on open standards, can be one effective technology of integrating transaction process of collaborative commerce. Charles Abrams, Gartner Group expert thinks that "Web Services have made a good development on those facets, including tools, infrastructure architecture, technique, standard. But, research on business model is short^[4]." So, based on Web Services, collaborative commerce model will be constructed and transaction process integration of collaborative commerce will be researched.

2. WEB SERVICES ARCHITECTURE

A Web service is an interface that describes a collection of operations that are network accessible through standardized XML messaging. A Web service is described using a standard, formal XML notion, called its service description. It covers all the details necessary to interact with the service, including message formats (that detail the operations), transport protocols and location^[5].

The Web Services architecture is based upon the interactions between three roles: service provider, service

registry and service requestor. The interactions involve the publish, find and bind operations. Together, these roles and operations act upon the Web Services artifacts: the Web service software module and its description. Figure1 illustrates these operations, the components providing them and their interactions^[5]

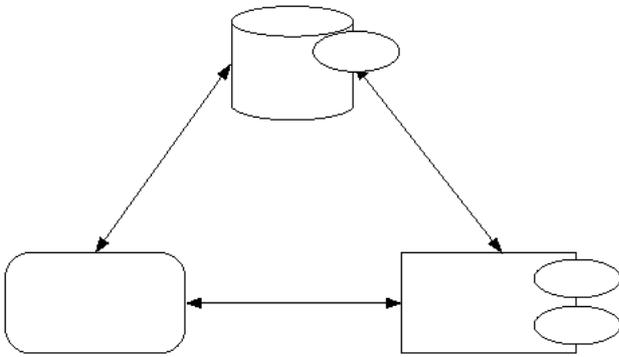


Figure1 Web Service architecture

In a typical scenario, a service provider hosts a network-accessible software module (an implementation of a Web service). The service provider defines a service description for the Web service and publishes it to a service requestor or service registry. The service requestor uses a find operation to retrieve the service description locally or from the service registry and uses the service description to bind with the service provider and invoke or interact with the web service implementation. Service provider and service requestor roles are logical constructs and a service can exhibit characteristics of both.

3. COLLABORATIVE COMMERCE MODEL BASED ON WEB SERVICES

3.1 Building Collaborative Commerce Model Based on Web Services

In web-enabled collaborative commerce and e-market mentioned above, only the buyer and seller interact with each other. However, it doesn't include the seller's interaction with the transportation company that will pick up and deliver the goods. This transaction also relies on the buyer and seller having an agreement to do business before the electronic transaction occurs^[6]. Obviously, the transaction process is not complete, and doesn't include value-added service.

Based on Web Services Service Oriented Architecture, collaborative commerce model was constructed as figure 2. This model expands the transaction boundary and incorporates all parties at the time of the transaction including third party logistics, third party storage, bank and etc. Thus, it can integrate all transaction process and all transaction data can flow freely between different systems. As a result, isolated information islands between different entities are eliminated. And also, it

provides accurate fulfillment information, dynamic trade agreement, and increased revenue recognition. Accurate fulfillment is a critical requirement for direct materials procurement. There's no tolerance in this market for out-of-stock conditions or backorders. The ability to dynamically trade with sellers will open up opportunities for e-hubs to emerge. These hubs can manage complex interactions for many businesses. Finally, directly incorporating creditors into the process will ensure faster revenue recognition for the seller and create a new financial market for underwriting and insuring these transactions^[6].

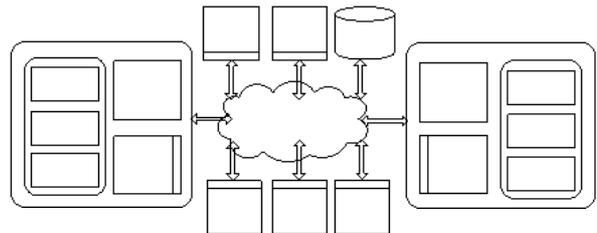


Figure2 Collaborative commerce model

Collaborative commerce model includes three kinds of entities:

- o Direct transaction entities, including buyer and seller. Anyone can be the service provider, also can be the service requestor. For example, manufacturer can publish order service to UDDI registry. At the same times, it also can request the third party logistics service.
- o Third party entities: including storage, logistics, bank and etc. The third party entities mainly are the role of services provider and provide services for business transaction.
- o UDDI Registry. It provides functions of publishing and finding of service.

3.2 Transaction Process of Collaborative Commerce Based on Web Services

Here on background of TV industry, we analyze how to integrate transaction process of collaborative commerce by Web Services. Supposed that there are five business entities included in the whole transaction process, TV distributor (buyer), TV manufacturer(seller), third party storage, third party logistics and bank. TV manufacturer published order service (OrderTV) to UDDI registry. Third party storage provides product query service (ProductQuery) to check product amount. Third party logistics provides service of transporting product (ProductTransport). Bank provides service of payment (AccountTransfer).

According to market forecasting, TV distributor plans to order one thousand of TVs. The whole transaction process of collaborative commerce is described as following:

- o TV distributor looks up order service(OrderTV) that TV manufacturer published in UDDI registry. There

Service
Requestor

Service
Description
Registry
Find
WSDL
UDDI
Bind
SOAP

are some open source software of UDDI registry, such SOAPUDDI and JUDDI. We can use find_business(), find_service(), find_tModel() and etc functions to get concrete information such as TV manufacturer, service list, bindingTemplate of every service, accessPoint of services and WSDL OverviewURL.

- ò According to returned WSDL document URL, TV distributor can use client tools to produce client proxy class, and use this proxy class invoke order service(OrderTV) of TV manufacturer. Here, we can use the tool of Axis, org.apache.axis.wsdl.Wsd12java, to produce client proxy class.

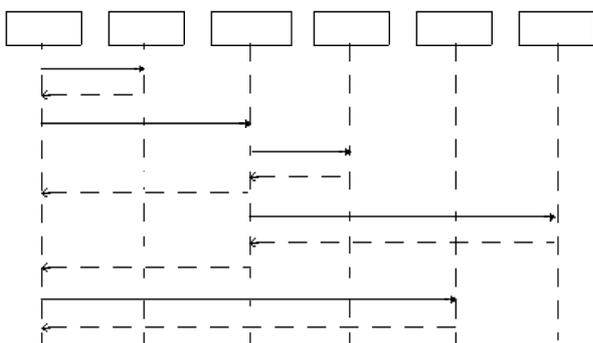


Figure3 Transaction process of collaborative commerce

- ò TV manufacturer invoke the product query service(ProductQuery) of third party storage to check the amount of TV in storage.
- ò According to the returned amount, TV manufacturer decided to accept order or reject.
- ò If accept order of distributor, TV manufacturer invokes transportation service(ProductTransport) of the third party logistics.
- ò After receive confirmation information of the third party logistics, TV manufacturer returns the state of order to TV distributor.
- ò The third party logistics delivers to TV distributor, order is fulfilled.
- ò TV distributor invoke service of payment (AccountTransfer) to give payment for TV, and the whole transaction is complete.

From above transaction process, we can see that every entity doesn't log on web site to fill in web form. Through direct interaction between systems, we integrate transaction processes of collaborative commerce and data of transaction seamlessly flow in and out of different system of business entities.

3.3 Implementation Way of Conversation Mechanism for Collaborative Commerce

In the ordering TV transaction process above, Web services of different business entities are included, and maybe the whole transaction last several days, even several weeks. So we should make mechanism of

conversation to keep track of execution process of whole transaction.

We can implement it by adding conversation state information to SOAP Header of buyer and seller. Conversation state information was inserted into SOAP Header by Service requestor's conversation state process module, and service responsor's conversation state process module use the information of services requestor. Concretely, we can use TPP/TPA of ebXML or cnXML.

The following is a simple demo of requestor's message, ConversationState SOAP Header item defines message exchange protocol of two sides of transaction, AgreementID element is ID number of protocol. ConversationID element is used to identify concurrent conversations under the same conversation agreement.

Conversation message demo of service requestor in SOAP message as following:

```
<env:Envelope xmlns:env="http://www.w3.org/2003/05/soap-envelope">
  <env:Header>
    <n:ConversationState xmlns:n="http://www.dlut.org/conversation">
      <n:AgreementId>
        uuid:0236223-8688-266b-f891-9dfdfj3j6jf
      </n:AgreementId>
      <n:conversationID>
        uuid:1323212-46bf-36fh-2323-dy1q2w5124
      </n:conversationID >
    </n: ConversationState >
  </env:Header>
  <env:Body>
    .....
  </env:Body>
</env:Envelope>
```

Message response uses ConversationState SOAP Header item which is produced by message requestor, so they are identical. During the life cycle of the whole transaction conversation, values of AgreementId and ConversationID don't be changed, and they are included in SOAP request message and response message.

4. CONCLUSION

Web Services are loosely coupled, based on open standard, and provide dynamically discovery transaction object. It becomes an effective technology in transaction process integration of collaborative. Deeply research on Web Service will promote rapid development of collaborative commerce.

REFERENCES

[1] Song, Liang, "Collaborative Commerce", <http://www.amteam.org/docs/bpwebsite.asp>, 2003.

Service Registry

Manufacturer

Third party storage

find manufacturer and services

overview URL and etc.

invoke service of ordering TV

- [2]Mike Clark, Peter Fletcher. *Web Services Business Strategies and Architectures*. Springer-Verlag New York Inc, pp234-238, 2002.
- [3]Xu Boyi, Jiang Lihong, "Implementation of collaborative commerce based on cnXML", *Journal of HeFei University of Technology*, vol. 26, pp695-699, 2003.
- [4]Charles Abrams, "Will Web Services Build Collaborative Commerce", http://www4.gartner.com/teleconferences/asset_47051.jsp, 2003.
- [5]Heather Kreger, "Web Services Conceptual Architecture", http://www-900.ibm.com/developerWorks/cn/webservices/ws-wsca/part1/index_eng.shtml, 2001.
- [6]JP Morgenthal, "Moving Toward Collaborative Commerce", *EAI Journal*, No. 5, pp25-26, 2001.

Research on Modular Business Processes

Weihui Dai¹, Weidong Zhao²

¹ School of Management, Fudan University, Shanghai 200433, China

² Software School, Fudan University, Shanghai 200433, China

¹whdai@vip.sina.com,²wdzhao@fudan.edu.cn

ABSTRACT

Business processes of traditional functional organizations are hard to respond to the rapidly changing market in the information era, so modular organizations are given attentions. In this paper, business processes in modular organizations based on the commitments between competence modules are analyzed and process reengineering of the Haier Company is used to illustrate the characteristics of modular organization processes.

Keywords: modular organizations, process, competence, commitment

1. INTRODUCTION

It is hard to predict market needs in the era of electronic business. With unpredictable change, organizations are expected to focus on not only the efficiency of processes, but also the flexibility and adaptability. Traditional functional organizations with large-scale production organized under the predictable requirements are product-driven, and concentrate on the business efficiency. The coordination among departments is made by management. In doing so, business processes are fragmented and rigid, being clumsy in responding to the competitive market. How to meet customers' needs rapidly becomes the challenge the sense-and-respond Organizations face (Stephan, 1999). By sense-and-response Organizations we mean that are customer centered, and assembled by numerous competence modules. The coordination among the competence modules is customer-driven by sensing and responding market change in complicated environments. Relative customer relations are formed by particular commitments between competence modules, which are managed and tracked by the commitment management system in order to integrate the ability and responsibility of them.

In this paper, we discuss how to apply the commitments between competence modules to analyze processes in sense-and-respond organizations, and a case study is illustrated.

2. PROCESS AS A COMPETENCY

Integrating organization elements such as information, material and human resources, a business process can transform certain inputs into outputs, which actually reflect one kind of competence that the process itself is a competency carrier. A value chain involves a few processes in the organization operations in order to create value for customers needing the cooperation of different departments such as procurement, manufacturing, sales, customer service and human resources, which form the multi-dimension of the

competence and present the modularity of the process. On the other side, the hierarchy of the process makes the competence module hierarchical. Several low-level complementary competence modules perform high-level processes, so the hierarchy and complementarity of processes represent the structure of organizations. Different processes involve different competent modules; the way in which they combine is also different. The relations of the main body of the competence module - the team or individuals who play one role and whose cooperation determines the performance of processes to a great extent - reflect the organizational form. Process change management such as BPR emphasizes breaking the bounds between functional departments to reduce the loss brought by the cooperation and intercourse or to build up the competence of one single module making the work team or individual responsible for more steps of the whole process.

From the view of the source of organization competitive advantages, it's important for an organization to possess more advantageous and heterogeneous resources than other organizations, but it can't create true value if the resources aren't transformed into the products or service customers need by organization units that add values (Colli, 2001). In essence, the real source of organizational competitive advantages is to create value for customers through the inimitable coordination between organizational competence modules and the effective use of the resources. The transformation needs the cooperation of interrelated competences, which make up a process. The operation, in fact, creates value for customs through the activities and cooperation of organizational units from different hierarchies. The competence of the organizational units is the ability to create values for the organization with resources allotted. The competence is delaminated, from individuals, departments, to teams or virtual organizations across departments or even organizations. The low-level competence is the basis of the high-level one; meanwhile the high-level competence takes on the integration effect because of the cooperation of the

low-level competences. From the low-level to high-level, the scope of integrations and degree of added values increase fast, causing more difficulties in organizational coordination. Organizational unit competence depends on not only specific resources and expertise, but also the structure of the organization, the way of organization units cooperate with. However, it will bring the negative effect if the units don't work well together. From the discussion above, a process is a harmonious value-adding carrier because it assembles various units to build the core competence of an organization. Of course, the harmonious environment will not appear automatically, and good organization structure, surveillance and adjustment of organization performance are needed. Herein, resources are the basis of the competence; organization units are the main body of the competence

The competitive strategy theory suggests that the competence is the embodiment of a set of business processes, which are well comprehensible from the view of business strategy. If an organization want to gain competitive advantage, it's necessary to transform its core business processes into the inimitable strategy implementation and provide super-value for customers. The essence of the transformation is to assemble different competence units into an inseparable whole across functions and even organizational boundary driven by market opportunities, which need close connection among interrelated processes. To do so, the adjustment in the organizational structure, strategic investment in IT and control of top management are necessary to guarantee the success of the transformation. It seems that the transformation is to determine and integrate necessary business processes to respond to customers, but it is not easy to change processes into strategic competence. Among the key success factors, the inventory replenishment process of Wal-Mart, for example, makes it an excellent retailer, which connects its suppliers, distribution centers and numerous POS in superstores via satellite communication systems and Internet, so that each actor in the supply chain can acquire necessary information in time, thereby the sale process of suppliers, distribution process of distribution centers and the order process of the POS conjoin closely.

3. THE PROCESSES OF MODULAR ORGANIZATIONS

With increasingly intense competition, it's vital for organizations to sense and respond to ever-changing environments quickly. As for this, the research has been done on adaptive organizations. In this paper, we discuss modular organizations based on the adaptive organization. Modular organizations are different from traditional ones, centering on customers and consist of competence modules. More importantly, these competence modules can be combined dynamically to build up stronger competence modules. Each

competence module owns necessary qualification and resources to play a certain role. Modular organizations are, therefore, highly adaptive, which are characterized by reengineering, agility and flexibility. Organizations can deploy the competence modules and their relations to satisfy customers so that the structure of modular organizations adapt to rapid reengineering of organization processes.

Customers are the reason and goal for the existence of competence modules, and market changes attribute to the reengineering of the competence modules. In modular organizations, the relationship between competence modules is customer-performer, which can request or provide service. Here, customers are a relative concept, referring to either external consumers (receivers of organization output) or internal customers.

The value of products or service provided can be evaluated by customers as the foundation of rewards. In process analyses, role-based cost management is taken to compare cost/return and discover the weak points in processes based on the value that each competence module creates for organizations. Besides, the contents and results of service are prescribed by commitments, which are one kind of formal or informal contract between competence modules. The commitments between the modules can be adjusted as a result of uncertainty of customers' expectations, which is achieved by continuous negotiations of the two. For instance, in the BPR of China Unicom Limited Heilongjiang Branch, operation units, as a customer of maintenance department, give the maintenance request and checkup standards. Then, the maintenance department will perform the maintenance and how well it work resorts to the operation units. In order to improve the efficiency of processes, commitments between competence modules can be managed by workflow management systems.

The adaptability of modular organizations depends largely on the smooth communication between competence modules, which needs an open atmosphere and active interactions, breaks the fencing of work dividing of functions and makes all parties participate in decision-making of organization planning.

Each module is authorized in modular organizations according to its capacity and responsibility. Decentralization and coordination are sometimes hard to compromise. How to coordinate competence modules to form an organic process system without strict command and control is a challenge to modular organizations. Generally speaking, common goals and effective coordination mechanisms between competence modules determine the way an organization runs as a whole. New competences will appear when the complementary modules are assembled organically, which avoids the isolation of traditional function departments and reduces the high cost of coordination. End consumers can't be satisfactory until commitments in every part of

processes are fulfilled. Therefore, commitment management is needed by all participants. On the other hand, how to capture external consumers to deploy competence modules and coordinate these modules is a problem.

From the view of point of commitments, roles of employees and work team etc. are identified and their contributions in the organization development are also easy to understand. Normative and authentic commitments simplify coordination management of modular organizations. In order to guarantee the cooperation among competence modules, commonly accepted protocols have to be designed. For instance, the commitment management of communication-based workflow models consists of four phases: definition, negotiation, execution and evaluation (Gregory, 2001). Commitments between competence modules are complex, speech and act theory, which is the basis of communication-based workflow models, are only simple abstraction of reality.

Modular organization processes are initiated from identifying customers' needs and then required competence modules and their commitments are designed. Herein competence modules are the combination of work teams of horizontal organizations. Actually, process management circulates the following four phrases: sensation, analysis, decision-making and execution. Change information in business environments is handled in the first two phases to provide the basis for subsequent reactions. To emphasize, competence modules can be taken by other organizations or even competitors (commitments between them may be more complicated). The commitment between competence modules is not to explain how results are accomplished, but to detail what the responsibilities are because the competence modules are automatic and, they can plan for their sub goals. The uncertainty of external customers may lead to the obscurity of the commitments, so the commitments between external customers and competence modules have to be adjusted repetitively. So do commitments between in-house modules of processes. Usually, commitments taken by a competence module are dual corresponding to its dual roles in a process: both a customer who requests service from the front role and a performer, who provides service for the next competence module in the process.

A process is, in essence, to transform external consumers' expectations into internal customers' requests, via the customer-performer chain of competence modules. When external consumers need new service, adjustments will be made along the chain. As a system, process modules must be well coordinated:

- Decompose process goals, and identify roles played and responsibilities taken on by competence modules. Every competence module should recognize that the

organization can satisfy external consumers and it can be rewarded only if it can satisfy its own "customer". Competence modules are designed to realize goals of organization processes, which is a top-to-bottom process of decomposing the goals

- Competence modules are customer-centered to establish the customer-performer chain through trust mechanism and commitment management. This actually introduces market mechanism to organization inside in order to improve the performance of competence modules, and to integrate external consumers with internal competence modules, even competitors into a whole for faster response.

- Competence modules can change commitments or reengineer. Changes from outside drive the commitments adjustment one-by-one along the customer-performer chain. On the other hand, information technology plays an important role in the establishment and maintenance of commitments among competence modules.

4. CASE STUDY: BPR OF HAIER GROUP

Linked by the market chain, the BPR of Haier encourages the initiative of employees and improvement of customer satisfaction. By the market chain we mean that market mechanism is introduced into organizations to prevent the coordination cost from exceeding the efficiency increase in labor division when organizations expand. That means that market is used as a coordination mechanism in Haier. The market chain turns internal relationships among processes, working procedures and employees into equal market relationships. Through the relationships, external market orders can be transformed into a series of internal "orders", forming a business process, which centers on orders with positions cooperating with each other and adjusting automatically. Customer-performer chains exist in organizational units. Whether commitments among units are fulfilled determine their revenues directly. The method above is essentially to build a benefit adjustment mechanism between internal organization units that reduces internal coordination costs. In doing so, the market chain transfers end consumers' requirements to every organizational unit, which facilitates information flow and feedbacks, making every organization unit aware of the competitive pressure and focusing on customers. In conclusion, the business processes of Haier are typical of a modular organization.

The market chain of Haier decomposes end consumers' needs into closely related customer-performer chains connected by commitments (market) as a whole. The market chain has overcome the embarrassment of isolated departments and no one being responsible for the whole process. Moreover, the market chain of Haier not only integrates internal organizational units through commitments, but also brings the suppliers and

customers together.

systems thinking.

5.SYSTEM THINKING

From the discussion above, it's concluded that each competence module in modular organizations plays a definite role in a process, and they can make their own plan independently. The process model takes a look on a customer-performer chain whose spots are competence modules. Can the goal of the whole process be achieved if every role has "satisfied" its own customers? The question is noticeable. A good example is the beer game mentioned in Peter Senge's *The Fifth Discipline* (Peter, 1990). This story is about the "game" among retailers, wholesalers and manufacturers in a beer supply chain. The three roles all attempt to meet customers' needs to gain more profit, but they are constrained by the limited rationality and have not considered upstream suppliers and downstream wholesalers, who are lack of communication and information sharing. As a result, when demands fluctuate, orders will be magnified along with the supply chain. This is the so-called bullwhip effect. This illustrates the necessity to alter the mental models of competence modules and strengthen the

ACKNOWLEDGEMENT

This paper is supported by the National Natural Science Foundation of China (No.70301004, No.70102004).

REFERENCES

- [1] Stephan, H Haeckel. *Adaptive Enterprise: Creating and Leading Sense-and-Respond Organizations*, New York: Harvard Business School Press, 1999
- [2] Colli, David J. *Company strategy*. Beijing China Renmin University Press, New York: Harvard Business School Press, 2001
- [3] egory Mentzas, Christos, Stylianos Kavadias. *Modeling business process with workflow systems: an evaluation of alternative approaches*. International Journal of Information Management, No. 21, pp123-135, 2001
- [4] Peter M. Senge. *The fifth discipline: the art and practice of the learning organization*, Shanghai: Shanghai San Lien Book Store, 1990

Research on Multi-Agent Based Information Management Platform for Enterprise Business Cooperation

Haigang Li

School of Management, Shanghai Jiao Tong University, Shanghai 200052 China
lihg@sjtu.edu.cn

ABSTRACT

The Enterprise Business Cooperation Mode is analyzed, as a result, an multi-agent based business cooperation framework is presented. The agent configuration in the framework is explained respectively, and a case of business process reengineering with multi-agent is discussed to prove the feasibility of the framework.

Key words: business cooperation; information integration; multi-agent technique; business process reengineering

1. INTRODUCTION

Looking back to the operation process of the well-known enterprises at home and abroad, it's easy to get a thought of enterprise' development in the future: internet provides more choice to the customer, customer's loyalty decides enterprise' survival and development in the age of internet economy, the enterprises will be sure to turn to pursuing the maximum of customer's satisfaction from simply pursuing the maximum of profit. In order to fulfill this goal, manufacture enterprise should make order-form-centered strategic cooperation with supplier, realizing interactivity, equity, win-win. Its core is to realize business cooperation including the information sharing and business cooperation of all the enterprises in the supply chain.

With the development of manufacturing industry mode and information technique especially, the enterprises are faced with a great amount of different distributing information resource. However, in the customer-oriented, rapidly changing and vehemently competitive market at present, enterprise's surviving room is tending to a state of "upper-flow" and "chaos"^[1]. The information system in Enterprise Business Cooperation Mode presents a characteristic of intermixing huge system. Traditional calculation mode can't deal with the complicated tasks in function and in structure. The appearance of CSCW (computer supported cooperative work) provides a better method to solve the problem, while Agent and MAS (multi-agent system) is the base of CSCW. Starting with analyzing the technique means, management notion and method adopted in enterprise's operation, business cooperation mode is discussed in aspect of system engineering, and combined with the Agent theory, an integrated framework suitable for business cooperation mode is constructed, providing a new approach for the study on Enterprise Integration System.

2. ANALYSIS OF ENTERPRISES' OPERATION ENVOROMENT

The worldwide market economy is rapidly developing

and the competition among enterprises is becoming more and more vehement. Traditional manufacturing industry including process enterprise operation mode has changed ultimately, from traditionally long-time, orderly manufacturing in manufacture plan organization to temporarily short-time, random manufacturing in order contract organization; from traditionally manufacturing characterized in large volume, single type, normative to temporarily manufacturing in small volume, various types, non-normative, from traditionally emphasizing on guaranteeing the carrying out of the manufacture techniques during the manufacture operation control to contemporary gaining the maximum economic benefit under the condition that the manufacture resources such as the material ,energy, human resource, capital and storage are limited. Therefore, many research organizations and scholars at home and abroad are studying in the field of E-commerce supply chain. Some advanced enterprises are trying to optimize the product's design and manufacture by using PDM (Product Data Management) and CPC (Cooperative Product Commerce) , reduce the internal cost by using ERP (Enterprise Resource Planning) , improve the whole supply chain's efficiency by using SCM (Supply Chain Management) , broaden the market and confirm the customer's loyalty by using CRM (Customer Relationship Management) , and operate from electron purchasing to solving problems in supply chain.

With the further studying of information technique and E-Business, the emphasis of supply chain's construction has been transferred from simply stressing on exchange to the whole transfer of business body (enterprise) and business activities in the network environment. To those enterprises that have adopted business cooperation mode, the application of "cooperation-trade" will bring the return of income and profit. The development of "cooperation-trade" will bring up highly agile dummy enterprises and bring change to some business models, such as material, the design and manufacture of products ^[2].

3. ENTERPRISES BUSINESS COOPERATION MODE

3.1 The Thought of Business Cooperation Mode

Business cooperation means not only the internal departments of the enterprise but also the associates including supplier, distributor, retailer and customers should be tied closely to exploit and meet the market's demand together, forming enterprise's dynamic union and cooperation. With unified plan and unified data mode, all the supply chain node enterprises make the product's cooperatively exploiting and material's cooperatively purchase, manufacture, distribution and delivery under the operation of unified plan. In a word, the cooperation among enterprises can be defined in three layers, one lies on operation layer, such as the purchase, promotion and order, etc, the second one lies on tactics layer, including the task-base forepart cooperation that work flow's upper-flow and down-flow enterprises make, and the third lies on strategic layer, that is, tight cooperation that work flow's upper and down. Enterprises start on phase of seeking for task. Enterprises involved the cooperation try their best to take advantage of the combination to form strong competition in addition to making use of their own advantage.

As a new subject, "cooperation" has been widely used as a crossing subject to study the common essential characteristics among different things. The notion of cooperation can be explained by simple mathematic symbols. The combination of any kind of product and market will have influence on enterprise's whole profit level. Suppose the price of some product is ¥ S , the expense on manpower, material, management and so on is ¥ O , and the investment on research and development, tools, workshop, inventory and etc. is ¥ I . The investment income per year on P_1 can be expressed as follows:

$$ROI = \frac{S_1}{O_1 - I_1} \quad (1)$$

As to other products P_2, P_3, \dots, P_n , the same expression can be got. If there's no relativity among products, the income of enterprise can be expressed as follows:

$$S_T = S_1 + S_2 + L + S_n \quad (2)$$

The whole operation cost and investment can be expressed as follows.

$$O_T = O_1 + O_2 + L + O_n \quad (3)$$

$$I_T = I_1 + I_2 + L + I_n \quad (4)$$

Therefore, the whole investment efficiency of enterprise is,

$$(ROI)_T = \frac{S_T - O_T}{I_T} \quad (5)$$

As long as the income, operation cost and investment of different products are irrelevant, all of the above relations will exist^[3]. Most enterprises have scale benefit. To a big company whose income is equal to the sum of

several small enterprises, its operation cost may be less than the sum of the small enterprises', or its total investment is less than the sum of theirs, which can be expressed as follows:

$$S_s = S_t; S_s \leq S_T; I_s \leq I_T; \quad (6)$$

In which, index s represents an enterprise, index t represents the aggregation of several single small enterprises. Therefore, when the income is equal, the investment income of the company is higher than that of the aggregation of several single small enterprises, that is:

$$(ROI)_s > (ROI)_t \quad (7)$$

When the total investment amount is equal, we can get the similar conclusion:

$$S_s \geq S_T; O_s \geq O_T; I_s = I_T \quad (8)$$

So we can know, manufacturing cooperatively can create higher income and/or lower operation cost. Taking advantage of cooperation can achieve all kinds of operating goal. If an enterprise has the ability to make use of the combination of products and market to increase the cooperation effect, the competition means offered will be very amplitude.

3.2 The Signification of Information Management Platform for Enterprise Business Cooperation

The information management platform for enterprises business cooperation integrates all the application and data into one platform, and supply the unitary user interface, which make enterprises build the information platform quickly between business to business, and business to its own employees. This platform is a comprehensive implementation mode of electronic business, and also a web based application system, which makes enterprises delivers all kinds of information stored inside or outside to employee, customer, suppliers and the firm comates, as a result, the information accessed by them is customization. The characteristics of the information platform can be summarized as the following, using unitary information access canal form which user can access their need information in order to optimize business operation and heighten productivity of company; using uninterrupted services through network and secure mechanism to make user access information anywhere and anytime; using strong content management capability to support almost all structure and non-structure data, identify the data from variant of data bases and collect and dispose all kinds of document; using individuation application services to set and supply the customization data and application such as enterprises portal in order to heighten the working efficiency of employee, intensify the affinity and attractive force; using integration with existing system to make the existing data and application be integrated together; using strong expansibility to adapt to new readjustment and change of enterprises, so to meet the need of business readjustment and expanding; using infallible and secure indemnity to guarantee the privacy and integrity of data; using collaboration and share to make user access and supply information for special

organization including individual.

3.3 Construct the Framework of Business Cooperation Model based on MAS technique

With introduction of mutual communication, distributed self-government agent, advanced enterprises cooperation business can be regarded as a heterogeneous multi-agent system which integrates all kinds of manufacturing action such as order, design, manufacturing and marketing together. This heterogeneous multi-agent is a polyphyletic cooperative system, in which each agent

team works for the same aims. We build the interactive enterprises cooperative business information management platform using object-oriented methods and MAS technology of artificial intelligence (see figure 1). From view of function realization, Agent is classified into two kinds, transaction agent and intra agent, the former take charge of information and communication between the key company in enterprises cooperative supply chain and the other node companies, while the latter is responsible for the information interflow and data commutation of different department within a enterprise.

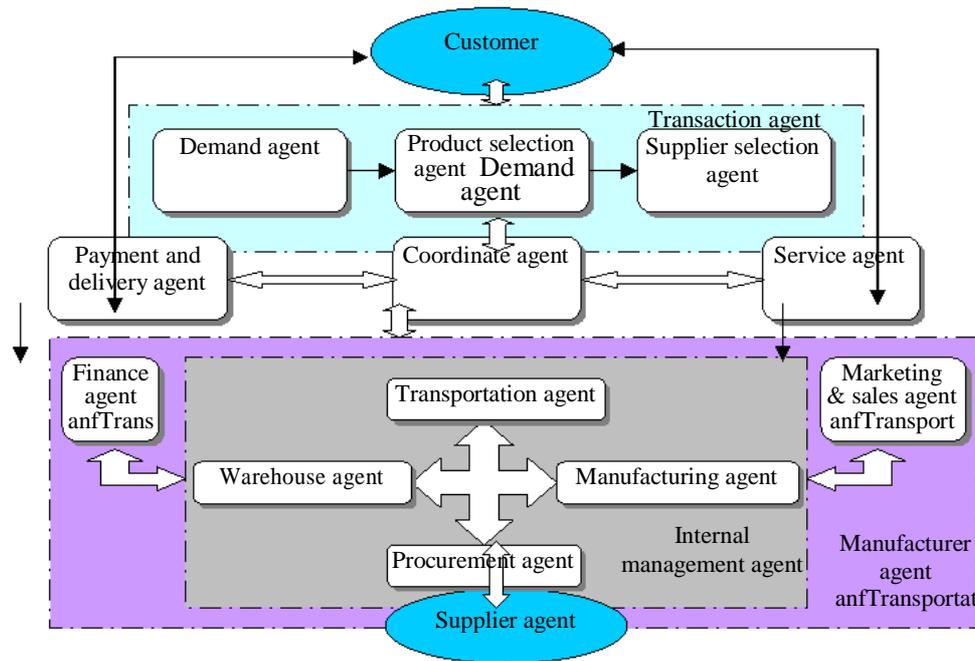


Fig.1 Enterprises business cooperative information management platform

4. CSASE STUDY

4.1 Design and implementation of MAS

Agent has the ability of manipulating multi-dimension information real time, impacting and communicating each other, which adapts to requirement of manipulation real time and concurrent for software development. On the other hand, development of internet technique make it possible for each agent unite to resolve problem cooperatively by teamwork in internet, with space distributed intelligence, information and technical resource. An integrated agent unite includes information model, resource model, organization model and cooperation model. Agents between upper layer and subordinate layer coordinate besides those in same layer. I.e., the original equal relationship is combined with the upper layer and subordinate layer relationship to form hierarchical relationship. The whole agents comprise an association in which agents are divided into several programmer teams according to task scope of agent. In a programmer team, agent is divided into three types

including management agent, function agent and service agent. The three type of agent lies in deferent layer and forms an agent tree, in which type and hierarchical relationship defines the status and power of the agent. Management and function agent comprise the key layer of programmer team, and all the agent association comprise the forest showed in figure2.

At different stage of design and development including view building and process flow model building, and agent model design and information support flat building of implementation stage, different modeling tool system and information integration techniques is needed. At requirement analysis stage there are all kinds of modeling tool software for function view and information view developed by IDEF0 and IDEF1 standard^[4], e.g., CASE and ARIS At design stage object modeling tool based on agent can be adopted such as ABE of IBM Co. Ltd. At implementation stage practical execution system is built in light of process flow model, the key to which is designing and packaging agent, building the information integrated flat for the object and

agents.

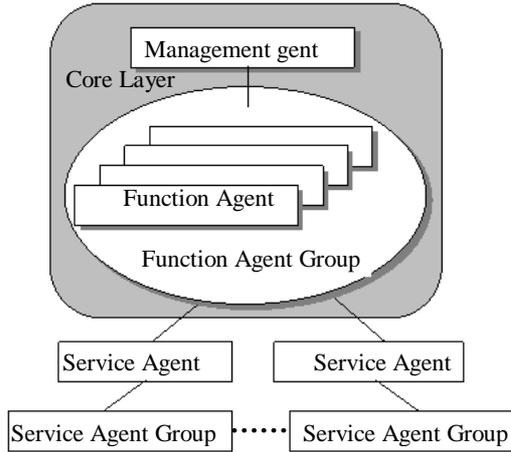


Fig. 2 Programmer organization model based on multi-agents

For the aim of inter-manipulation, multi-agent system needs support of standard and operation system. One is KQML standard, a language and protocol for exchanging information and knowledge. It supports the information share among agents as a language standard and information management protocol; meanwhile it is used as an interface among multi-agent system for resolving protocol problem as application programming language. Another is CORBA standard, provides service from abstract distributed object model and coincidence depiction of language, shields isomer of operation flats and data resource, to make different application transplant and manipulate each other. The third is COM/DCOM standard, a binary unit model defined for inter-manipulation of units, is a basement unit model which is irrespective of programming language and suitable for application and system development of units.

4.2. Application in sales workflow of cooperative business information platform

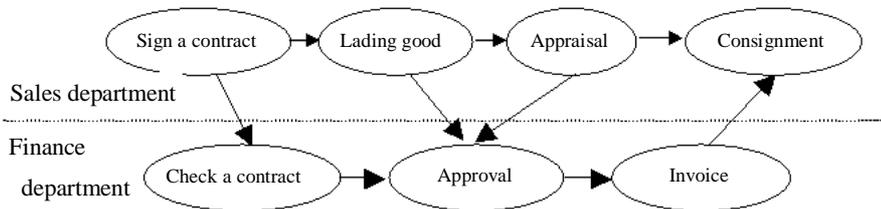


Fig.3 Sales and consignment process of a company

5. CONCLUSION

Today in the world, the enterprise is up against new challenging from the tendency of economy globalization and the enterprises development idea of centering on customer service. It is seems very necessary and important for corporations to reengineering the business process, integrated the forthcoming application, data and

The process flow modeling based on Agent expresses all the process flow activities and their relations by Agent model. Agent's characteristics of autonomy, agility, interaction and cooperation facility enterprise to restructure and reengineer the manufacture and operating process and to combine tightly with the information supporting flat during the implementation. The topological structure and incidence relation of agent Network in the Cooperation Layer are the reflection of the model of enterprise business disposal process. Every agent in cooperation layer corresponds coarse grain business disposal process in real life and has certain stability and atomicity. According to the value chain transitive theory, there is some un-value added action in figure 3. Suppose each business disposal spot corresponds to an agent in cooperation layer, in order to support business dynamic adjustment, every knowledge base of business disposal agent should include the knowledge on dynamic business disposal. For instance, if the following rules are added to the knowledge base of consignment agent, the consignment flow can be decides according to the priority of the contract.

```

If contract data agent.priority (contract number)>xx
then
consignment agent.consignment (bill of lading number)
else
If invoice agent.completion (contract number) =Ture
then consignment agent.consignment (bill of lading
number) else consignment agent.hanging
End if
End if

```

Every agent has the unitive expression about its own information and the interface information of the related agent (which composes the topological structure of agent network and reflects the enterprise business model), invoke of methods between different agents can be fulfilled naturally and conveniently, as a result, it can support the dynamic enterprise model strongly.

the performance of enterprises integrated flexibility.

REFERENCES

- [1]. Athans M. "Command and control theory: a challenge M. to control science. *IEEE Trans on AC*, vol. 32, No. 4, pp286-293, 1987.
- [2]. Shyama V Ramani. "Technology cooperation between firms of developed and less-developed countries". *Economics Letters*, No. 68, pp203-209, 2000.
- [3]. B. Kadar et al. "An Object-Oriented Framework for Developing Distributed Manufacturing Architectures". *Journal of Intelligent Manufacturing*. Vol. 9, No. 2, pp 173-179, 1998.
- [4]. Chan Yuliu, etc. "The Current Design Analysis Method of IDEF0 and IDEFIX". *Beijing: Electron Industry Publishing House*, 1991.

Study of cooperative product business support platform for new product development

Haigang Li

School of Management, Shanghai Jiao Tong University, Shanghai 200052, China
lihg@sjtu.edu.cn

ABSTRACT

The development phases of e-business is described firstly, then the new product development process in cooperative product business is presented in detail, the cooperative product business support platform including technical platform and information platform is discussed finally.

Key words: new product development; e-business; cooperative product business; platform

1. INTRODUCTION

With the more and more fierily of market competition, enterprise is obliged to adopt new administration operation mechanism and new business model, the key to this is electronic business. The kernel meaning can be presented like this, firstly, new business model should be built and the existing business model should be integrated on the basis of the Internet techniques platform; secondly, organization structure of enterprises should be reengineering and improved systematically by the technique of business process reengineering for the purpose of enhancement of enterprises operation; thirdly, relationship among manufacturer, supplier and customer should be redefined and extended by the technique of

supply chain management and customer relationship management; fourthly, internal and external resource of enterprises should be integrated, and attention to internal value chain should be changed to external one by the method of business process integration; finally, digital operation and decision should be fulfilled by the method of business intelligence.

Product is the key to manufacturing operation of enterprise. For manufacturing industry, cooperative business based on Internet is the senior type of electronic business, which includes electronic commerce and collaborative product commerce. Figure 1 gives us the development process of electronic business.

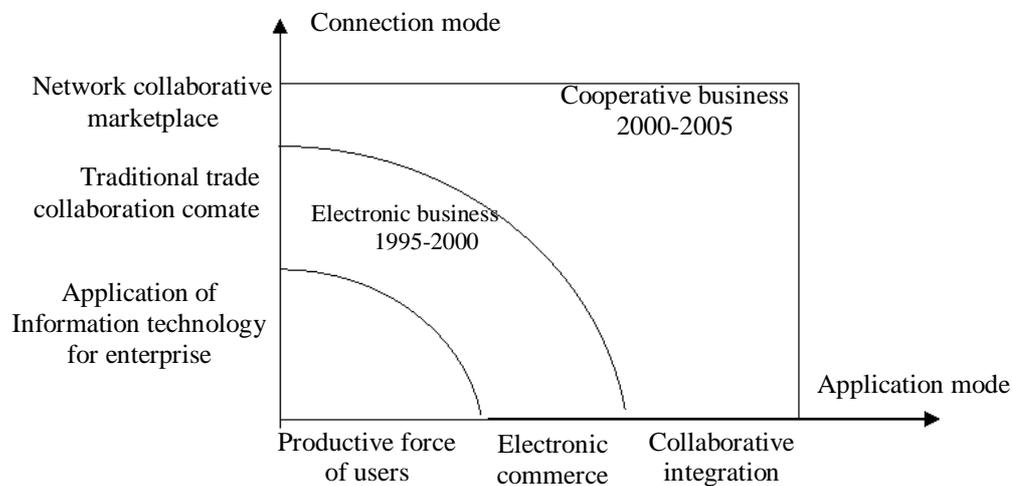


Fig. 1 The development phases of e-business

2. NEW PRODUCT DEVELOPMENT PROCESS IN COOPREATIVE BUSINESS MODE

With the coming of electronic business times, more and more manufacturing enterprises is thinking about how to integrate the company itself, suppliers and customer together and fulfill the collaborative development and sales of product by the network. As a result,

manufacturer and supplier will face the market together in electronic business times, and form the firm dynamic alliance by the method of collaborative product business. In this mode, the firm alliance subject builds its own collaborative business platform and fulfills the development and collaborative management of digitization prototype inside the kernel firm.

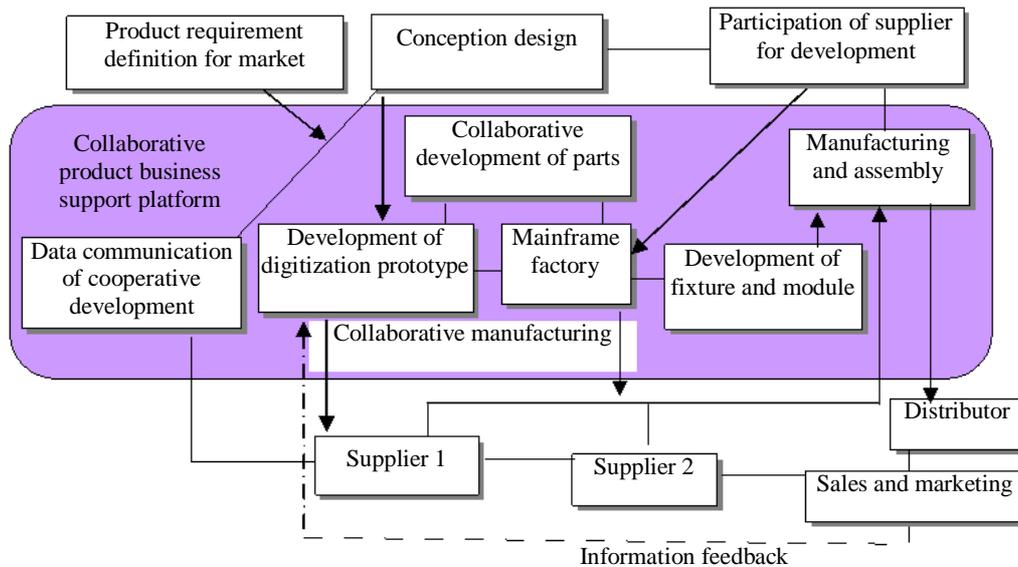


Fig. 2 A new product development process under cooperative business environment

On this platform, digital prototype development and collaborative management is fulfilled within the key firm, extends forwards to customer end through product requirement definition of customer, customer relationship management for market projection, and backward to maintenance, subassembly supply, distributor and customer relationship management for afterservice. In the medial product development link, the platform extends to supply chain to assure business to business model. Figure 2 gives us the new product development process under cooperative product business environment. Firstly, the function, performance, market object price, time putting on sale, supply chain and other strategy is defined by the key firm so as to map out the new product development plan according market

requirement and customer value added factor. Secondly, development project team of new product is built including the technique department inside the firm and supplier so as to develop the whole or parts of the new product with support of cooperative product business support platform,. During the process, the information of design, manufacturing, test and feedback for new product is exchanged and shared. Finally, when the new product has been cast to market, value added service after sale is extended to distributor and eventual customer though the cooperative product business support platform, including of supply chain management of parts, fault diagnosis and maintenance, and also the new product innovation information can be acquired form market and customer as the reference of the next new product development cycle.

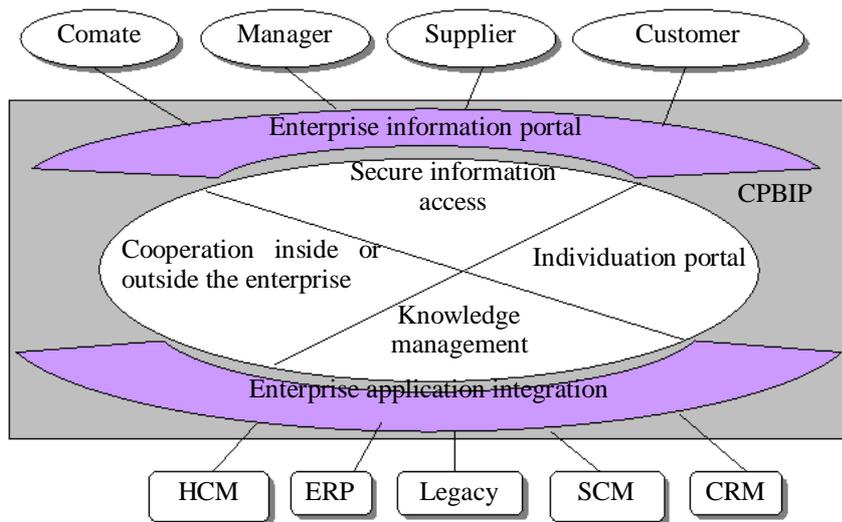


Fig. 3 the architecture of CPBIP

3. COLLABORATIVE PRODUCT BUSINESS SUPPORT PLATFORM FOR NEW PRODUCT DEVELOPMENT

The collaborative product business support platform

(abbreviated as CPBSP) for new product development is constituted by two main parts, one is CPBIP, a information support environment formed by Enterprise information portal, enterprise application integration,

knowledge management system and some secure information access techniques. the other is CPBTP, a technology support environment formed by series of product design and manufacturing techniques for virtual digitization prototype mode such as computer aided designing (CAD) /computer aided engineering (CAE)/ computer aided manufacturing (CAM) , virtual product data management (VPDM) / product data management (PDM).

The cooperative product business information platform uses the new Internet technology to fulfill the different business coordination inside the whole supply chain or across the supply chain, and change the enterprise business operation mode. The architecture of CPBIP is showed as figure 3.

Figure 4 gives us the architecture of the CPBTP. In organization, CPBSP is a network system formed by

persons from different tasks team such as product design, product analysis, manufacturing, simulation and support. In process, CPBSP supply the digital and virtual support for new product development, the new product is developed in method of concurrent engineering. In formality, the product under the support of CPBSP is a digital model actually which can be observed, analyzed, communicated and regenerated in demand, thus the design, analysis, producibility, maintainability and marketability of the product can be coordinated at the same time. In benefit and effectiveness, one product model is used in development such that the non-value added time is shorted and problem can be found and solved in time. Communication and coordination between enterprises in different place can be executed better so every worker in participation of the new product development can acquire his needed information in demand.

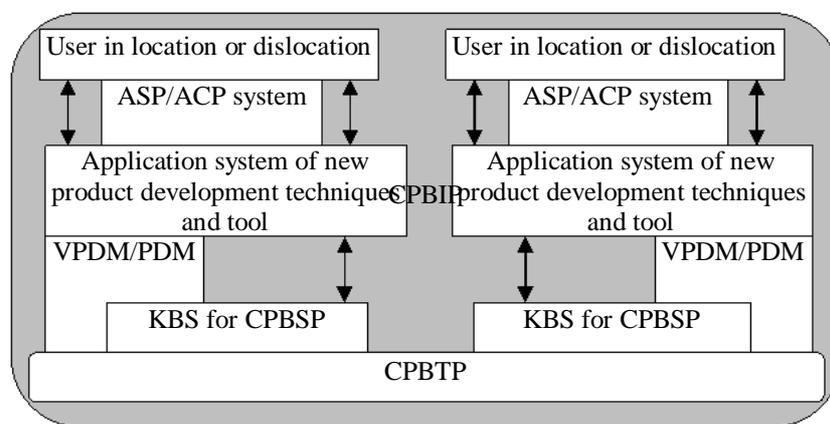


Fig. 4 The collaborative product business support platform

4. CONCLUSION

As an advanced technique, the cooperative product support platform is used in the new product development process by enterprise to response to customer rapidly and flexibly, so the new product can be cast to market with prescriptive time, cost and quality. Through the CPBSP, different users within the key firm can use CPBTP to support the development of new product, and cooperative strategy comate can use either its own CPBTP or CPBIP to communicate with the member firms in the supply chain. The new product development under the cooperative business environment is becoming a strong measure to carry out the strategy translation for manufacturing industry and is paid high attention by the enterprise. The study of cooperative product business support platform is significant to the new product development of the future.

REFERENCE

- [1]. Conn. Stamford. Gartner Group identifies "C-Commerce" supply chain moevement: An Emerging trend in Collaborative Web Communities[EB/OL], <http://citebm.business.uiuc.edu/emba/references/c-commerce.html>, 1999-08-06.
- [2]. Phillips C, Meeker M. The B2B internet report-collaborative commerce[EB/OL], <http://www.morganstanley.com/institutional/techrese/arch/pdfs/b2bpla.pdf>, 2000-04.
- [3]. IBM. B2B collaborative Commerce with the sametime, Quickplace and WebSphere Commerce suite[EB-OL], <http://www.redbooks.ibm.com/redbooks/SG246218.html>, 2001-08.

Supporting Collaborative Communication in a Multi-layer Meta-process Model for Evolutionary Shared Workflows

Damien Bright¹, Gerald Quirchmayr^{1,2}, Roger Tagg¹

¹School of Computer & Information Science, Univ. of South Australia, Mawson Lakes SA 5095, Australia

²Institut für Informatik und Wirtschaftsinformatik, Univ. Wien, Liebiggasse 4/3-4 A-1010 Vienna, Austria
{Damien.Bright,Gerald.Quirchmayr,Roger.Tagg}@unisa.edu.au

ABSTRACT

A key planning activity within a Virtual Enterprise (VE) is to establish agreed inter-organizational processes. This activity, or meta-process, has to allow for gradual evolution of the VE processes and for a multi layer development from informal business agreements to precise workflows. To support this meta-process, a collaborative electronic whiteboard supported by a tuplespace is proposed. The whiteboard supports a mixed graphical and text interface, with support for keeping track of the changes made. The participating organizations upload workflow definitions from their own IT systems into the tuplespace. Workflows are then discussed, modified and evolved before being downloaded again and mapped to the partners' individual systems.

Keywords: Virtual Enterprises, B2B Collaboration, Inter-enterprise workflows, Electronic Whiteboards, Tuplespaces

1. INTRODUCTION

In an earlier paper [1], two of the current authors addressed the issue of IT support for developing shared workflows in Virtual Enterprises (VEs). VEs are characterized as formalized, but not permanent, alliances negotiated between autonomous organizations, where the goal is to present seamless business services to customers. Members are encouraged to agree upon and then enact processes that follow overarching guidelines evolved by the VE. These processes may evolve during negotiations to establish the VE or during the lifetime of the VE as conditions of the alliance are changed. Processes may eventually be discontinued when the alliance is dissolved. Negotiating efficient cross organizational processes can be of critical importance to the business value of the VE.

In this paper we define a *shared workflow* as a business process that two or more organizational entities agree to follow. A shared workflow is related to the private workflows of the individual organizations, but may only contain a partial view of them; each participating organization may not want elements of its internal processes to be externally observable. The term *evolutionary* means that shared workflows have not all been settled in advance, but need to be negotiated incrementally in step with the evolving business agreements between the organizations. A *meta-process* is a procedure – or process - whose purpose is to establish a business process that will be enacted for operational business cases. The term *multi-layer* derives from the earlier paper, in which several layers in the meta-process were identified, the first three being a) informal discussion b) agreeing a business process model and c) agreeing an executable workflow system.

The VE, or the subset of members involved in each particular business collaboration, needs to reach a final agreement on one or more inter-organizational processes that represent the way they agree to inter-operate. However each individual organization will usually also need to adjust its private processes to fit in with the new collaboration. For example, it may need to provide additional facilities for notifying partner organizations of progress, for responding to requests for status reports, or for dealing with exceptions such as non-response from the partner.

The question is - how best can IT be used to help the meta-process? Our intuition is that even if the VE has developed guidelines, the meta-process can never be as strictly predefined as with typical production workflows. But it is equally unrealistic to assume that the meta-process is totally informal. One possible intermediate paradigm is a negotiation model, but typical negotiation models focus on a limited number of measured utility parameters such as price, delivery time etc. Our task is to find something equivalent to a negotiation model that will allow the parties to make proposals, counter-proposals and agreements on a whole range of issues such as the nature of tasks, the order they should occur in, the data that should be provided, the persons that are authorized to commit to a decision, the rules governing exceptional situations - and so on.

This paper proposes a way to support the collaborative meta-process of evolving a set of workflows that will coordinate the activities of a VE. The collaborative activity is assumed to take place over the internet in the online equivalent of a meeting room [2], and that communication can be abstracted to the exchange of a series of events related to the collaboration. It includes the use of graphics exchange protocols for real time.

However our approach aims to avoid explicit dependence on any particular process model diagram or language standard, other than that the model can be represented in XML.

In this paper, section 2 summarizes the related work in business process modelling in B2B, business contract architectures, inter-enterprise workflows, evolutionary workflows, whiteboard/blackboard models, and coordination models using tuplespaces - with related software developments. Section 3 describes the architecture and basic model of our proposed approach. Section 4 discusses a possible way of implementing our approach with available software. Section 5 presents our conclusions and directions for further work.

2. RELATED WORK

ebXML [3] has been proposed as a standard approach to B2B e-commerce. It envisages that, at the level of the structure and sequence of messages to be passed, each organization proposes a Coordination Protocol Profile (CPP); the organizations then negotiate an agreed protocol (CPA). However ebXML has not yet recommended a meta-process for agreeing a CPA – use of an ebXML-familiar broker (consultant) is assumed.

Automatic and formal approaches for aligning business processes from different organizations have been proposed by Mahleko [4] and Grossmann et al [5]. However the orientation here is to find an *integrated* process, rather than the shared sub-process that represents only what happens *between* the organizations.

Negotiation of a shared process and negotiation of a contract are very closely linked. The contract will specify the duties of each party in both physical and informational terms. A description of our work in conjunction with the Business Contracts Architecture of Milosevic et al is given in [6] and [7].

The CrossFlow project [8], which is based on services-based cross-organizational workflow (SCW), looks at contracts as the key element in defining inter-enterprise workflows. Other important issues in CrossFlow include reliability, trust and service management (see also [9]).

Collaboration involving workflows in virtual enterprises has been addressed by Ludwig and Whittingham [10] and Wynen et al [11].

Analyses of alternative architectures for inter-organizational workflow have been made by van der Aalst [12]. Schulz and Milosevic [13] have carried out a good analysis of cross-organizational process architectures, primarily oriented to B2B E-commerce. Wing, Liu and Colomb [14] introduced the notion of incremental trading, implying a loosely-coupled VE in which the members gradually get to know more about each other. Action Workflow [15] is also relevant as a well-established workflow approach (and software

product) that is based on a pattern of exchanges of information along the lines of offer and acceptance. This closely mirrors the situation in many shared inter-organizational workflows.

In terms of the process of evolving workflows through collaboration and negotiation, the most notable work was done on the ORCHESTRA Esprit project [16].

In the pre-electronic age, the people negotiating the shared process would most likely meet in a room with a white (or black!) board and interactively agree the processes and data standards. The web based equivalent of this environment is a virtual meeting room [2] with support for a shared digital whiteboard. Shared whiteboards are used by many desktop conferencing systems, including Microsoft's NetMeeting. A state of the whiteboard is determined by its initial state and a sequence of events that are incrementally applied to it. Some of the issues involved in transmission of data streams used by common shared whiteboard tools are described in [17].

The idea of a tuplespace has its origins in coordination languages such as Linda [18]. A tuple is a vector of fields. Users of a tuplespace can post and retrieve tuples using coordination primitives [19]. A number of extensions of the Linda primitives have been introduced; some of these offer *reactive programming* [20] functionality. This can support, for example, triggering of user notifications or program calls on an event such as the insertion of tuples into a tuplespace.

An XML document can be stored in a tuple field value (or values). The ability to match on nodes in the XML document using some form of query language such as XQL is supported by tuplespace implementations such as TSpaces [21] and XMLSpaces [22]. TSpaces also includes reactive programming primitives, including the ability to register with the server for notifications based on whether a tuple that matches a specific template is written to the server by any client.

The most relevant work to date in combining Linda-based workspaces with workflows is described by Tolksdorf [23].

3. ARCHITECTURE AND BASIC MODEL

3.1 Multi-layer model

The previous paper [1] saw the evolution of business processes for a VE as a multiple set of layers, each of which would need to be agreed as part of the meta-process. Since the system level implementation of the processes will generally involve application layer protocols, it is useful to regard the layers as an extension of the standard seven layer ISO/OSI (Open Systems Interconnection) reference model. The extra layers introduced were as follows:

Table 1: Layers of a meta-process as an extension above the ISO/OSI communications profile, from [1]

Layer	Nature of interaction
11	Informal discussions of terms of agreement, procedures and rules
10	Tailoring of proposed or potential process patterns by cooperative and interactive use of whiteboard technology.
9	Agreeing formal and/or executable models incorporating process and data, using shared data and workflow models
8	Agreeing domain-specific models for terminology, and for task and data transfer
7a	Agreeing standard message structures (XML schemas etc) for carrying transferred processes, tasks and data
7b	Coordinating requests between computer software systems across data communication facilities

At layer 11 the negotiation is between senior business managers. They may introduce a mediator, or may resort to a tried and tested set of choreographed exchanges to perform the negotiation, similar to an auction with set rules. The emphasis is on the broad principles of the collaboration. This includes determining which parts of the process are to be under joint control and which are a purely local responsibility. The data being passed at this layer is mainly informal. If it is recorded at all, it is likely to be free text based, e.g. emails with optional attachments. These text messages can be supported in our system as message based tuples.

At layer 10, more detailed negotiation takes place (involving business process analysts, lawyers, accountants etc) to work out potential process patterns. The subject matter includes not only the basic control flow in the process model, but also such things as required data, authorized roles, time limits, action to be taken on exceptions – in fact anything that could appear in a legal clause. At this level, the parties may not be using compatible – or indeed any – formal diagramming notations. However whiteboard-style negotiation will be easier if some simple diagramming is used.

At layer 9 the meta-process would work with formal diagrams representing data structures and process patterns. The process diagram needs to show the process dependencies in the shared workflow. Since the organizations may use different modelling conventions, a standard format is used on the whiteboard, which is mapped to the individual organizations' models using WfXML. However it should be noted that many workflow systems include more than just control flow, and may employ more than one type of diagram for capturing the total design of a workflow pattern.

At layer 8 the meta-process is limited to agreement on whether or not to use a certain domain-specific standard, e.g. RosettaNet [24] for the IT electronics industry. We are not concerned with what happens below layer 8. Agreement on standard message structures and basic request-response communications

is made between the IT managers of the organizations. In many cases, the decision will be made when each organization chooses its B2B software.

3.2 Motivating example

We consider a simple case where two organizations currently both offer a set of business services that allow customers to place single and multiple orders, enquire about prices and availability etc. Because they operate in related markets, the organizations wish to introduce a number of shared processes to streamline the service they can offer the customers. The idea is that each organization will offer the other organization's products as well as its own, but would need to check availability and prices across the organizational boundary. As a result, a) each partner's service interface would need to be adjusted and b) a shared process would need to be started up.

A tuplespace is also useful for supporting dynamic groups of users where participants can join and leave at any time. A participant who joins an ongoing session can retrieve, by matching against a set of tuples, existing data exchanged by other participants since the start of the session. This is useful in the support of the changing membership of alliances that is often common in VEs.

A tuplespace is also useful for supporting dynamic groups of users where participants can join and leave at any time. A participant who joins an ongoing session can retrieve, by matching against a set of tuples, existing data exchanged by other participants since the start of the session. This is useful in the support of the changing membership of alliances that is often common in VEs.

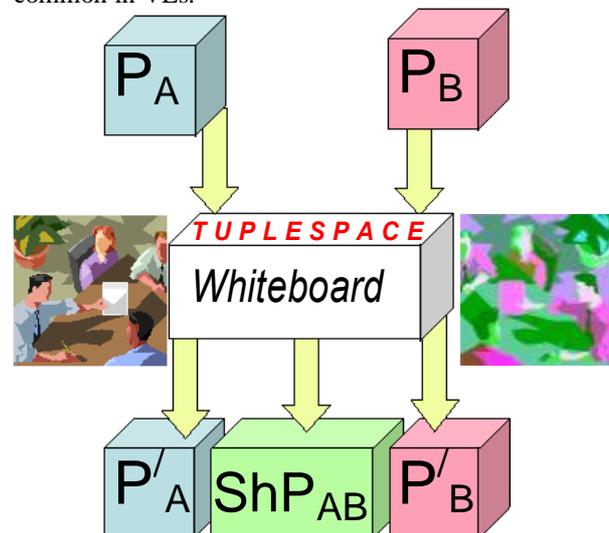


Figure 1: Overview of the whiteboard-assisted meta-process

The meta-process changes the situation from two separate processes P_A and P_B to a shared process ShP_{AB} flanked by two private processes P'_A and P'_B , as shown in Figure 1 above. ShP_{AB} may be more than a simple choreography if brokers or other intermediaries are

involved. Figure 2 explains what is contained within the processes P_A etc. The interface (at the front) can be with either a human user or a program (e.g. part of an end customer's purchase order system). The control flow determines the hierarchy, sequence, branching, parallelism and iteration of a set of activities (human or computer). Secondary perspectives cover all other information necessary to fully define the process.

3.2 Representation of the model as a tuplespace

A tuplespace approach has advantages for the type of collaboration scenario we wish to support because it offers functionality similar to both active database and pattern-based messaging. It can also support situations where the relationship between data items is dynamic and transitory rather than following rigid data models.

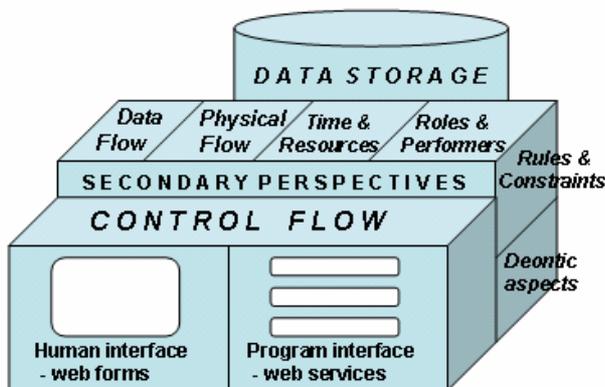


Figure 2: Contents of a customer process

A tuplespace is also useful for supporting dynamic groups of users where participants can join and leave at any time. A participant who joins an ongoing session can retrieve, by matching against a set of tuples, existing data exchanged by other participants since the start of the session. This is useful in the support of the changing membership of alliances that is often common in VEs.

The workflow definitions being evolved could be in many different forms, such as a) a business process modelling tool; b) a workflow management system or c) workflow facilities within an enterprise resource planning (ERP) system such as SAP [25]. The key to being able to correlate processes is the existence of reliable mappings between the workflow definitions and the XML data represented in the tuples in the tuplespace. We use Wf-XML [26] as a unifying format, although this will not in general be sufficient to cover all the process semantics. We propose to supplement Wf-XML with the use of a *constraint document* encoded in OCL (Object Constraint Language) [27]. Beyond this there may still be a need to hold other tuple-based data structures as required.

In our approach we first translate the definitions of the existing workflows into Wf-XML, which provides the initial state of the tuplespace. The combined workflows are then presented for discussion in a neutral graphical format, using swim lanes to distinguish the different

partners and intermediaries. Input and output messages are also shown on the graph, and can be expanded by clicking. In the discussion the order of actions may be changed; actions may be inserted or removed; or the constraint document modified to change triggers for certain actions, or to change the conditions for transitions between actions to occur. When a participant in the meta-process makes a change, we capture these changes as generic whiteboard events, by adding tuples to the tuplespace. In order to preserve the order of incremental state changes, we keep track of event sequence numbers.

4. IMPLEMENTATION APPROACH

A tuplespace implementation which matches our requirements for shared virtual data spaces on the internet is TSpaces [21] developed by IBM. A TSpaces server can support multiple tuplespaces instances and a client can locate a TSpaces server on the web and then attach to a named tuplespace instance hosted on the server. Once a client connection to a tuplespace is established the various supported TSpaces operations can be used to interact with tuples (read, write, take, update, delete etc). Access control is implemented at the tuplespace level so that each operation requested by a client to be performed on a particular tuplespace instance has a related list of AccessAttributes that need to be satisfied for the operation to be allowed. As part of the supported operations there are blocking commands where a client can use a WaitToRead or WaitToTake command. The TSpaces server implements a callback functionality which can be used to inform clients of the arrival of tuples matching an outstanding query. Clients can therefore setup notifications which can be used to keep track of a set of required tuples (e.g. client can wait for tuples required to perform an action).

In terms of XML support, the TSpaces server supports the use of an XML field in a tuple and creates a *tuple tree* for XML documents which mirrors the DOM (Document Object Model) tree which corresponds to the XML structure. A subset of the XQL query language is supported to allow path expression based matching on data within an XML document. IBM has also developed the TSpaces services suite, which helps to automate the development and management of web services using tuplespaces. It is interesting to note that XMLSpaces [22], which is a direct extension of TSpaces, has added new functionality to provide a more detailed incorporation of XML into tuple fields and XML support for the original Linda definition of the matching relation for tuples. XMLSpaces also extends TSpaces by offering distributed tuplespace support across multiple servers.

We are experimenting with Version 2.1.2 of the TSpaces server software running on a web server with an additional software layer that is used for transformations and connectivity with SAP Netweaver [28]. Netweaver is a Wf-XML compatible tool for

orchestrating both SAP and non-SAP services that is based on SAP Business Workflow. Examples of transformations needed by the system are the mapping of whiteboard protocols to coordination primitives and the use of XSL [29] for transformation between various XML document structures. A limited set of basic TSpaces tuple data structures have been defined to represent evolutionary workflows to be exchanged in the system:

```
WorkflowTuple = (PartnerRole, Id,
                TimeStamp, XMLDoc, [ActionTuple]+)
ActionTuple = (PartnerRole, Id, XMLDoc)
MessageTuple = (PartnerRoleTo,
                PartnerRoleFrom, Id, Subject, Text,
                XMLDoc)
```

where + represents a field which can be repeated multiple times, PartnerName identifies a particular business partner, Id is a unique number which can be used as a way to track sequences and set membership, and XMLDoc represents an XML document structure. TSpaces-supported coordination primitives can be used to interact with these tuples and to generate other temporary data tuples as needed, using reactive programming techniques.

5. CONCLUSION AND FUTURE WORK

Our proposed system supports collaborative activities for evolving workflows for a virtual enterprise, based upon (a) interactive use of a shared whiteboard (for tailoring potential process patterns based on graphical interaction) and (b) cooperative discussions of the terms of agreement, procedures and rules using a blackboard architecture (interaction based on message exchange). We have described how this can be achieved using tuplespaces, and are working on a prototype implementation using IBM's TSpaces.

Future work plans include 1) improved handling of the secondary perspectives, possibly incorporating a general ontology for process negotiation; 2) specialization of the whiteboard interface for the different layers (i.e. top management (layer 11), business analysts and lawyers (layer 10), and line managers and process specialists (layer 9); and 3) catering for volatile N-way negotiations including intermediaries as well as partners who join and leave VE agreements.

We also intend to continue working towards the presentation of a full prototype implementation.

REFERENCES

- [1] Tagg, R., Quirchmayr, G., "Towards an interconnection model for evolution of shared workflows in a virtual enterprise", *proceedings of 3rd Intl Conf. on Information Integration and Web-Based Applications and Services (iiWAS)*, Linz, Austria, September, 2001
- [2] Roseman, M., Greenberg, S., TeamRooms: Network Places for Collaboration, *proceedings of ACM CSCW'96 Conference*, ACM Press 1996
- [3] ebXML, "ebXML Deliverables web site", <http://www.ebxml.org/specs/> last accessed October 2004
- [4] Mahleko, B., "Efficient MatchMaking of Business Processes for B2B Process Integration", *proceedings of 10th Doctoral Consortium, Conference on Advanced Information Systems Engineering (CaiSE)*, Klagenfurt, Austria, 2003 <http://www.ipsi.fraunhofer.de/oasys/reports/ftp/pdf/P2003-05.pdf>
- [5] Grossmann, G., Schrefl, M., Stumptner, M., Ren, Y., "Behavior based integration of business processes by examining states and activities", *Technical Report*, University of South Australia, 2004
- [6] Quirchmayr G., Milosevic Z., Tagg R., Cole J., Kulkarni S., "Establishment of virtual enterprise contracts", *proceedings of DEXA'02 conference*, Aix-en-Provence, France, September 2002
- [7] Tagg, R., Milosevic, Z., Kulkarni, S., Gibson, S., "Supporting contract execution through recommended workflows", *proceedings of DEXA'04 conference*, Zaragoza, Spain, September 2004
- [8] Grefen P., Aberer K., Hoffner Y., Ludwig H., "CrossFlow: cross-organizational workflow management in dynamic virtual enterprises", *International Journal of Computer Systems Science & Engineering* Vol 15, No 5, pp 277-290, 2000
- [9] Klingemann, J., Wäsch, J., Aberer, K., Deriving service models in cross-organizational workflows, *proceedings of RIDE –Information technology for Virtual Enterprises (RIDE-VE'99)*, Sydney, Australia, March 1999.
- [10] Ludwig, H., Whittingham, K., "Virtual enterprise co-ordinator – agreement-driven gateways for cross-organisational workflow management", *proceedings of WACC'99 conference*, IEEE, February 1999
- [11] Wynen F., Perrin O., Bitcheva J., Godart C., "A model to support collaborative work in virtual enterprises", *proceedings of BPM'2003 conference*, Eindhoven, Holland, Springer LNCS 2678, pp. 104-119
- [12] van der Aalst, W., "Process-oriented architectures for electronic commerce and interorganizational workflow", *Information Systems Journal*, Pergamon/Elsevier, 2000
- [13] Schulz, K., Milosevic, Z., "Architecting cross-organisational B2B interactions", *proceedings of EDOC 2000 Conference*, IEEE, 2000
- [14] Wing, H., Liu, C., Colomb, R., "A bottom-up approach to distributed workflow", *proceedings of*

- Pacific Asia Conference on Information Systems (PACIS)*, Brisbane, Australia, 1997
- [15] Medina-Mora, R., Winograd, T., Flores, R., Flores, F., “The Action Workflow approach to workflow management technology”, *proceedings of ACM Conference on Computer Supported Cooperative Work (CSCW)*, 1992
- [16] Guimaraes, N., Pereira, A-P., Antunes, P., (1997) “Bridging workflow and collaboration tools”, *proceedings of 7th Mini EURO conference*, Bruges, Belgium, 1997
<http://www.di.fc.ul.pt/~paa/papers/minieuro-97.pdf>
- [17] Geyer, W., Weis, R., “A secure, accountable and collaborative whiteboard”, *proceedings of IDMS'98 conference*, Oslo, Norway, Spriner LNCS 1483, pp 3-14, 1998
- [18] Ahuja, S., Carriero, N., Gelernter, D., “Linda and friends”, *IEEE Computer*, 19 (8), pp 26-34
- [19] Andrade, L., Fiadeiro, J., “Coordination primitives for evolving event-based systems”, *proceedings of 1st Workshop on Evolutionary Formal Software Development*, Copenhagen, Denmark, pp 19-30, (2002)
- [20] Busi N., Rowstron A., Zavattaro G., “State and event-based reactive programming in shared dataspace”, *proceedings of Intl Conference on Coordination Models and Languages (Coordination'02)*, Springer LNCS pp 111-124 (2002)
- [21] Fontoura, M., Lehman, T., Nelson, E., Truong, T., Xiong, Y., (2003), “TSpaces services suite: automating the deployment and maintenance of web services”, *proceedings of 12th International World Wide Web Conference (WWW'2003)*, Budapest, Hungary, ACM Press, 2003
- [22] Tolksdorf R., Glaubitz D., “XMLSpaces for coordination in web-based systems”, *proceedings of 10th IEEE Intl Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, IEEE Computer Society Press, 2001
- [23] Tolksdorf, R., “Coordination Technology for Workflows on the Web: Workspaces”, *proceedings of 4th International Conference on Coordination Models and Languages (Coordination2000)*, Springer LNCS, pp 36–50, 2000
- [24] RosettaNet, “RosettaNet home page”, <http://www.rosettanet.org/RosettaNet/Rooms/DisplayPages/LayoutInitial> , last accessed October 2004
- [25] SAP, “SAP Solutions web index page”, <http://www.SAP.com/solutions> , last accessed October 2004
- [26] Workflow Management Coalition, “Workflow Standard – Interoperability: Wf-XML Binding”, *WfMC Document Number WfMC-TC-1023*, 2000
<http://www.wfmc.org/standards/docs/Wf-XML-1.0.pdf> , last accessed October 2004
- [27] Object Management Group (OMG), “UML 2.0 OCL (Object Constraint Language) Specification”, 2003, <http://www.omg.org/docs/ptc/03-10-14.pdf> , last accessed October 2004
- [28] Ferguson, R., “Great expectations for SAP's NetWeaver”, *EWeek Magazine*, April 19, 2004, <http://www.eweek.com/article2/0,1759,1569238,00.asp> last accessed October 2004
- [29] World Wide Web Consortium (W3C), “XSL Transformations (XSLT) Version 1.0”, 1999
<http://www.w3.org/TR/xslt> , last accessed October 2004.

Supporting Collaborative Communication in a Multi-layer Meta-process Model for Evolutionary Shared Workflows

Damien Bright¹, Gerald Quirchmayr^{1,2}, Roger Tagg¹

¹School of Computer & Information Science, Univ. of South Australia, Mawson Lakes SA 5095, Australia

²Institut für Informatik und Wirtschaftsinformatik, Univ. Wien, Liebiggasse 4/3-4 A-1010 Vienna, Austria
{Damien.Bright,Gerald.Quirchmayr,Roger.Tagg}@unisa.edu.au

ABSTRACT

A key planning activity within a Virtual Enterprise (VE) is to establish agreed inter-organizational processes. This activity, or meta-process, has to allow for gradual evolution of the VE processes and for a multi layer development from informal business agreements to precise workflows. To support this meta-process, a collaborative electronic whiteboard supported by a tuplespace is proposed. The whiteboard supports a mixed graphical and text interface, with support for keeping track of the changes made. The participating organizations upload workflow definitions from their own IT systems into the tuplespace. Workflows are then discussed, modified and evolved before being downloaded again and mapped to the partners' individual systems.

Keywords: Virtual Enterprises, B2B Collaboration, Inter-enterprise workflows, Electronic Whiteboards, Tuplespaces

1. INTRODUCTION

In an earlier paper [1], two of the current authors addressed the issue of IT support for developing shared workflows in Virtual Enterprises (VEs). VEs are characterized as formalized, but not permanent, alliances negotiated between autonomous organizations, where the goal is to present seamless business services to customers. Members are encouraged to agree upon and then enact processes that follow overarching guidelines evolved by the VE. These processes may evolve during negotiations to establish the VE or during the lifetime of the VE as conditions of the alliance are changed. Processes may eventually be discontinued when the alliance is dissolved. Negotiating efficient cross organizational processes can be of critical importance to the business value of the VE.

In this paper we define a *shared workflow* as a business process that two or more organizational entities agree to follow. A shared workflow is related to the private workflows of the individual organizations, but may only contain a partial view of them; each participating organization may not want elements of its internal processes to be externally observable. The term *evolutionary* means that shared workflows have not all been settled in advance, but need to be negotiated incrementally in step with the evolving business agreements between the organizations. A *meta-process* is a procedure – or process - whose purpose is to establish a business process that will be enacted for operational business cases. The term *multi-layer* derives from the earlier paper, in which several layers in the meta-process were identified, the first three being a) informal discussion b) agreeing a business process model and c) agreeing an executable workflow system.

The VE, or the subset of members involved in each particular business collaboration, needs to reach a final agreement on one or more inter-organizational processes that represent the way they agree to inter-operate. However each individual organization will usually also need to adjust its private processes to fit in with the new collaboration. For example, it may need to provide additional facilities for notifying partner organizations of progress, for responding to requests for status reports, or for dealing with exceptions such as non-response from the partner.

The question is - how best can IT be used to help the meta-process? Our intuition is that even if the VE has developed guidelines, the meta-process can never be as strictly predefined as with typical production workflows. But it is equally unrealistic to assume that the meta-process is totally informal. One possible intermediate paradigm is a negotiation model, but typical negotiation models focus on a limited number of measured utility parameters such as price, delivery time etc. Our task is to find something equivalent to a negotiation model that will allow the parties to make proposals, counter-proposals and agreements on a whole range of issues such as the nature of tasks, the order they should occur in, the data that should be provided, the persons that are authorized to commit to a decision, the rules governing exceptional situations - and so on.

This paper proposes a way to support the collaborative meta-process of evolving a set of workflows that will coordinate the activities of a VE. The collaborative activity is assumed to take place over the internet in the online equivalent of a meeting room [2], and that communication can be abstracted to the exchange of a series of events related to the collaboration. It includes the use of graphics exchange protocols for real time.

However our approach aims to avoid explicit dependence on any particular process model diagram or language standard, other than that the model can be represented in XML.

In this paper, section 2 summarizes the related work in business process modelling in B2B, business contract architectures, inter-enterprise workflows, evolutionary workflows, whiteboard/blackboard models, and coordination models using tuplespaces - with related software developments. Section 3 describes the architecture and basic model of our proposed approach. Section 4 discusses a possible way of implementing our approach with available software. Section 5 presents our conclusions and directions for further work.

2. RELATED WORK

ebXML [3] has been proposed as a standard approach to B2B e-commerce. It envisages that, at the level of the structure and sequence of messages to be passed, each organization proposes a Coordination Protocol Profile (CPP); the organizations then negotiate an agreed protocol (CPA). However ebXML has not yet recommended a meta-process for agreeing a CPA – use of an ebXML-familiar broker (consultant) is assumed.

Automatic and formal approaches for aligning business processes from different organizations have been proposed by Mahleko [4] and Grossmann et al [5]. However the orientation here is to find an *integrated* process, rather than the shared sub-process that represents only what happens *between* the organizations.

Negotiation of a shared process and negotiation of a contract are very closely linked. The contract will specify the duties of each party in both physical and informational terms. A description of our work in conjunction with the Business Contracts Architecture of Milosevic et al is given in [6] and [7].

The CrossFlow project [8], which is based on services-based cross-organizational workflow (SCW), looks at contracts as the key element in defining inter-enterprise workflows. Other important issues in CrossFlow include reliability, trust and service management (see also [9]).

Collaboration involving workflows in virtual enterprises has been addressed by Ludwig and Whittingham [10] and Wynen et al [11].

Analyses of alternative architectures for inter-organizational workflow have been made by van der Aalst [12]. Schulz and Milosevic [13] have carried out a good analysis of cross-organizational process architectures, primarily oriented to B2B E-commerce. Wing, Liu and Colomb [14] introduced the notion of incremental trading, implying a loosely-coupled VE in which the members gradually get to know more about each other. Action Workflow [15] is also relevant as a well-established workflow approach (and software

product) that is based on a pattern of exchanges of information along the lines of offer and acceptance. This closely mirrors the situation in many shared inter-organizational workflows.

In terms of the process of evolving workflows through collaboration and negotiation, the most notable work was done on the ORCHESTRA Esprit project [16].

In the pre-electronic age, the people negotiating the shared process would most likely meet in a room with a white (or black!) board and interactively agree the processes and data standards. The web based equivalent of this environment is a virtual meeting room [2] with support for a shared digital whiteboard. Shared whiteboards are used by many desktop conferencing systems, including Microsoft's NetMeeting. A state of the whiteboard is determined by its initial state and a sequence of events that are incrementally applied to it. Some of the issues involved in transmission of data streams used by common shared whiteboard tools are described in [17].

The idea of a tuplespace has its origins in coordination languages such as Linda [18]. A tuple is a vector of fields. Users of a tuplespace can post and retrieve tuples using coordination primitives [19]. A number of extensions of the Linda primitives have been introduced; some of these offer *reactive programming* [20] functionality. This can support, for example, triggering of user notifications or program calls on an event such as the insertion of tuples into a tuplespace.

An XML document can be stored in a tuple field value (or values). The ability to match on nodes in the XML document using some form of query language such as XQL is supported by tuplespace implementations such as TSpaces [21] and XMLSpaces [22]. TSpaces also includes reactive programming primitives, including the ability to register with the server for notifications based on whether a tuple that matches a specific template is written to the server by any client.

The most relevant work to date in combining Linda-based workspaces with workflows is described by Tolksdorf [23].

3. ARCHITECTURE AND BASIC MODEL

3.1 Multi-layer model

The previous paper [1] saw the evolution of business processes for a VE as a multiple set of layers, each of which would need to be agreed as part of the meta-process. Since the system level implementation of the processes will generally involve application layer protocols, it is useful to regard the layers as an extension of the standard seven layer ISO/OSI (Open Systems Interconnection) reference model. The extra layers introduced were as follows:

Table 1: Layers of a meta-process as an extension above the ISO/OSI communications profile, from [1]

Layer	Nature of interaction
11	Informal discussions of terms of agreement, procedures and rules
10	Tailoring of proposed or potential process patterns by cooperative and interactive use of whiteboard technology.
9	Agreeing formal and/or executable models incorporating process and data, using shared data and workflow models
8	Agreeing domain-specific models for terminology, and for task and data transfer
7a	Agreeing standard message structures (XML schemas etc) for carrying transferred processes, tasks and data
7b	Coordinating requests between computer software systems across data communication facilities

At layer 11 the negotiation is between senior business managers. They may introduce a mediator, or may resort to a tried and tested set of choreographed exchanges to perform the negotiation, similar to an auction with set rules. The emphasis is on the broad principles of the collaboration. This includes determining which parts of the process are to be under joint control and which are a purely local responsibility. The data being passed at this layer is mainly informal. If it is recorded at all, it is likely to be free text based, e.g. emails with optional attachments. These text messages can be supported in our system as message based tuples.

At layer 10, more detailed negotiation takes place (involving business process analysts, lawyers, accountants etc) to work out potential process patterns. The subject matter includes not only the basic control flow in the process model, but also such things as required data, authorized roles, time limits, action to be taken on exceptions – in fact anything that could appear in a legal clause. At this level, the parties may not be using compatible – or indeed any – formal diagramming notations. However whiteboard-style negotiation will be easier if some simple diagramming is used.

At layer 9 the meta-process would work with formal diagrams representing data structures and process patterns. The process diagram needs to show the process dependencies in the shared workflow. Since the organizations may use different modelling conventions, a standard format is used on the whiteboard, which is mapped to the individual organizations' models using WfXML. However it should be noted that many workflow systems include more than just control flow, and may employ more than one type of diagram for capturing the total design of a workflow pattern.

At layer 8 the meta-process is limited to agreement on whether or not to use a certain domain-specific standard, e.g. RosettaNet [24] for the IT electronics industry. We are not concerned with what happens below layer 8. Agreement on standard message structures and basic request-response communications

is made between the IT managers of the organizations. In many cases, the decision will be made when each organization chooses its B2B software.

3.2 Motivating example

We consider a simple case where two organizations currently both offer a set of business services that allow customers to place single and multiple orders, enquire about prices and availability etc. Because they operate in related markets, the organizations wish to introduce a number of shared processes to streamline the service they can offer the customers. The idea is that each organization will offer the other organization's products as well as its own, but would need to check availability and prices across the organizational boundary. As a result, a) each partner's service interface would need to be adjusted and b) a shared process would need to be started up.

A tuplespace is also useful for supporting dynamic groups of users where participants can join and leave at any time. A participant who joins an ongoing session can retrieve, by matching against a set of tuples, existing data exchanged by other participants since the start of the session. This is useful in the support of the changing membership of alliances that is often common in VEs.

A tuplespace is also useful for supporting dynamic groups of users where participants can join and leave at any time. A participant who joins an ongoing session can retrieve, by matching against a set of tuples, existing data exchanged by other participants since the start of the session. This is useful in the support of the changing membership of alliances that is often common in VEs.

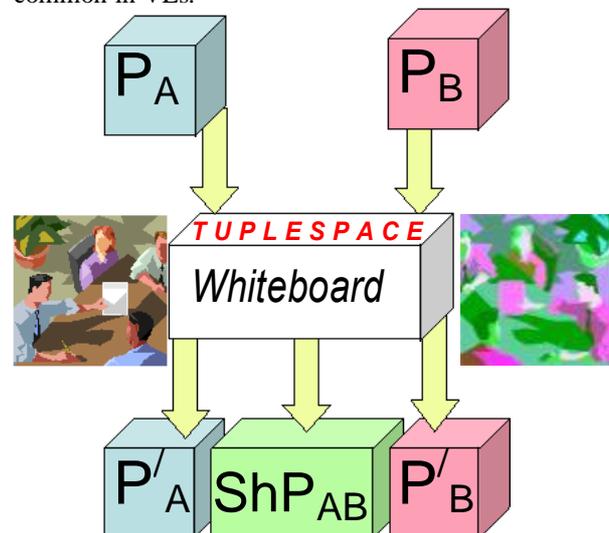


Figure 1: Overview of the whiteboard-assisted meta-process

The meta-process changes the situation from two separate processes P_A and P_B to a shared process ShP_{AB} flanked by two private processes P'_A and P'_B , as shown in Figure 1 above. ShP_{AB} may be more than a simple choreography if brokers or other intermediaries are

involved. Figure 2 explains what is contained within the processes P_A etc. The interface (at the front) can be with either a human user or a program (e.g. part of an end customer's purchase order system). The control flow determines the hierarchy, sequence, branching, parallelism and iteration of a set of activities (human or computer). Secondary perspectives cover all other information necessary to fully define the process.

3.2 Representation of the model as a tuplespace

A tuplespace approach has advantages for the type of collaboration scenario we wish to support because it offers functionality similar to both active database and pattern-based messaging. It can also support situations where the relationship between data items is dynamic and transitory rather than following rigid data models.

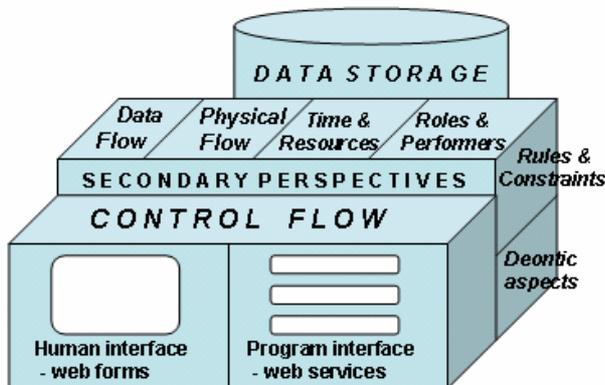


Figure 2: Contents of a customer process

A tuplespace is also useful for supporting dynamic groups of users where participants can join and leave at any time. A participant who joins an ongoing session can retrieve, by matching against a set of tuples, existing data exchanged by other participants since the start of the session. This is useful in the support of the changing membership of alliances that is often common in VEs.

The workflow definitions being evolved could be in many different forms, such as a) a business process modelling tool; b) a workflow management system or c) workflow facilities within an enterprise resource planning (ERP) system such as SAP [25]. The key to being able to correlate processes is the existence of reliable mappings between the workflow definitions and the XML data represented in the tuples in the tuplespace. We use Wf-XML [26] as a unifying format, although this will not in general be sufficient to cover all the process semantics. We propose to supplement Wf-XML with the use of a *constraint document* encoded in OCL (Object Constraint Language) [27]. Beyond this there may still be a need to hold other tuple-based data structures as required.

In our approach we first translate the definitions of the existing workflows into Wf-XML, which provides the initial state of the tuplespace. The combined workflows are then presented for discussion in a neutral graphical format, using swim lanes to distinguish the different

partners and intermediaries. Input and output messages are also shown on the graph, and can be expanded by clicking. In the discussion the order of actions may be changed; actions may be inserted or removed; or the constraint document modified to change triggers for certain actions, or to change the conditions for transitions between actions to occur. When a participant in the meta-process makes a change, we capture these changes as generic whiteboard events, by adding tuples to the tuplespace. In order to preserve the order of incremental state changes, we keep track of event sequence numbers.

4. IMPLEMENTATION APPROACH

A tuplespace implementation which matches our requirements for shared virtual data spaces on the internet is TSpaces [21] developed by IBM. A TSpaces server can support multiple tuplespaces instances and a client can locate a TSpaces server on the web and then attach to a named tuplespace instance hosted on the server. Once a client connection to a tuplespace is established the various supported TSpaces operations can be used to interact with tuples (read, write, take, update, delete etc). Access control is implemented at the tuplespace level so that each operation requested by a client to be performed on a particular tuplespace instance has a related list of AccessAttributes that need to be satisfied for the operation to be allowed. As part of the supported operations there are blocking commands where a client can use a WaitToRead or WaitToTake command. The TSpaces server implements a callback functionality which can be used to inform clients of the arrival of tuples matching an outstanding query. Clients can therefore setup notifications which can be used to keep track of a set of required tuples (e.g. client can wait for tuples required to perform an action).

In terms of XML support, the TSpaces server supports the use of an XML field in a tuple and creates a *tuple tree* for XML documents which mirrors the DOM (Document Object Model) tree which corresponds to the XML structure. A subset of the XQL query language is supported to allow path expression based matching on data within an XML document. IBM has also developed the TSpaces services suite, which helps to automate the development and management of web services using tuplespaces. It is interesting to note that XMLSpaces [22], which is a direct extension of TSpaces, has added new functionality to provide a more detailed incorporation of XML into tuple fields and XML support for the original Linda definition of the matching relation for tuples. XMLSpaces also extends TSpaces by offering distributed tuplespace support across multiple servers.

We are experimenting with Version 2.1.2 of the TSpaces server software running on a web server with an additional software layer that is used for transformations and connectivity with SAP Netweaver [28]. Netweaver is a Wf-XML compatible tool for

orchestrating both SAP and non-SAP services that is based on SAP Business Workflow. Examples of transformations needed by the system are the mapping of whiteboard protocols to coordination primitives and the use of XSL [29] for transformation between various XML document structures. A limited set of basic TSpaces tuple data structures have been defined to represent evolutionary workflows to be exchanged in the system:

```
WorkflowTuple = (PartnerRole, Id,
                TimeStamp, XMLDoc, [ActionTuple]+)
ActionTuple = (PartnerRole, Id, XMLDoc)
MessageTuple = (PartnerRoleTo,
                PartnerRoleFrom, Id, Subject, Text,
                XMLDoc)
```

where + represents a field which can be repeated multiple times, PartnerName identifies a particular business partner, Id is a unique number which can be used as a way to track sequences and set membership, and XMLDoc represents an XML document structure. TSpaces-supported coordination primitives can be used to interact with these tuples and to generate other temporary data tuples as needed, using reactive programming techniques.

5. CONCLUSION AND FUTURE WORK

Our proposed system supports collaborative activities for evolving workflows for a virtual enterprise, based upon (a) interactive use of a shared whiteboard (for tailoring potential process patterns based on graphical interaction) and (b) cooperative discussions of the terms of agreement, procedures and rules using a blackboard architecture (interaction based on message exchange). We have described how this can be achieved using tuplespaces, and are working on a prototype implementation using IBM's TSpaces.

Future work plans include 1) improved handling of the secondary perspectives, possibly incorporating a general ontology for process negotiation; 2) specialization of the whiteboard interface for the different layers (i.e. top management (layer 11), business analysts and lawyers (layer 10), and line managers and process specialists (layer 9)); and 3) catering for volatile N-way negotiations including intermediaries as well as partners who join and leave VE agreements.

We also intend to continue working towards the presentation of a full prototype implementation.

REFERENCES

- [1] Tagg, R., Quirchmayr, G., "Towards an interconnection model for evolution of shared workflows in a virtual enterprise", *proceedings of 3rd Intl Conf. on Information Integration and Web-Based Applications and Services (iiWAS)*, Linz, Austria, September, 2001
- [2] Roseman, M., Greenberg, S., TeamRooms: Network Places for Collaboration, *proceedings of ACM CSCW'96 Conference*, ACM Press 1996
- [3] ebXML, "ebXML Deliverables web site", <http://www.ebxml.org/specs/> last accessed October 2004
- [4] Mahleko, B., "Efficient MatchMaking of Business Processes for B2B Process Integration", *proceedings of 10th Doctoral Consortium, Conference on Advanced Information Systems Engineering (CaiSE)*, Klagenfurt, Austria, 2003 <http://www.ipsi.fraunhofer.de/oasys/reports/ftp/pdf/P2003-05.pdf>
- [5] Grossmann, G., Schrefl, M., Stumptner, M., Ren, Y., "Behavior based integration of business processes by examining states and activities", *Technical Report*, University of South Australia, 2004
- [6] Quirchmayr G., Milosevic Z., Tagg R., Cole J., Kulkarni S., "Establishment of virtual enterprise contracts", *proceedings of DEXA'02 conference*, Aix-en-Provence, France, September 2002
- [7] Tagg, R., Milosevic, Z., Kulkarni, S., Gibson, S., "Supporting contract execution through recommended workflows", *proceedings of DEXA'04 conference*, Zaragoza, Spain, September 2004
- [8] Grefen P., Aberer K., Hoffner Y., Ludwig H., "CrossFlow: cross-organizational workflow management in dynamic virtual enterprises", *International Journal of Computer Systems Science & Engineering* Vol 15, No 5, pp 277-290, 2000
- [9] Klingemann, J., Wäsch, J., Aberer, K., Deriving service models in cross-organizational workflows, *proceedings of RIDE –Information technology for Virtual Enterprises (RIDE-VE'99)*, Sydney, Australia, March 1999.
- [10] Ludwig, H., Whittingham, K., "Virtual enterprise co-ordinator – agreement-driven gateways for cross-organisational workflow management", *proceedings of WACC'99 conference*, IEEE, February 1999
- [11] Wynen F., Perrin O., Bitcheva J., Godart C., "A model to support collaborative work in virtual enterprises", *proceedings of BPM'2003 conference*, Eindhoven, Holland, Springer LNCS 2678, pp. 104-119
- [12] van der Aalst, W., "Process-oriented architectures for electronic commerce and interorganizational workflow", *Information Systems Journal*, Pergamon/Elsevier, 2000
- [13] Schulz, K., Milosevic, Z., "Architecting cross-organisational B2B interactions", *proceedings of EDOC 2000 Conference*, IEEE, 2000
- [14] Wing, H., Liu, C., Colomb, R., "A bottom-up approach to distributed workflow", *proceedings of*

- Pacific Asia Conference on Information Systems (PACIS)*, Brisbane, Australia, 1997
- [15] Medina-Mora, R., Winograd, T., Flores, R., Flores, F., “The Action Workflow approach to workflow management technology”, *proceedings of ACM Conference on Computer Supported Cooperative Work (CSCW)*, 1992
- [16] Guimaraes, N., Pereira, A-P., Antunes, P., (1997) “Bridging workflow and collaboration tools”, *proceedings of 7th Mini EURO conference*, Bruges, Belgium, 1997
<http://www.di.fc.ul.pt/~paa/papers/minieuro-97.pdf>
- [17] Geyer, W., Weis, R., “A secure, accountable and collaborative whiteboard”, *proceedings of IDMS'98 conference*, Oslo, Norway, Spriner LNCS 1483, pp 3-14, 1998
- [18] Ahuja, S., Carriero, N., Gelernter, D., “Linda and friends”, *IEEE Computer*, 19 (8), pp 26-34
- [19] Andrade, L., Fiadeiro, J., “Coordination primitives for evolving event-based systems”, *proceedings of 1st Workshop on Evolutionary Formal Software Development*, Copenhagen, Denmark, pp 19-30, (2002)
- [20] Busi N., Rowstron A., Zavattaro G., “State and event-based reactive programming in shared dataspace”, *proceedings of Intl Conference on Coordination Models and Languages (Coordination'02)*, Springer LNCS pp 111-124 (2002)
- [21] Fontoura, M., Lehman, T., Nelson, E., Truong, T., Xiong, Y., (2003), “TSpaces services suite: automating the deployment and maintenance of web services”, *proceedings of 12th International World Wide Web Conference (WWW'2003)*, Budapest, Hungary, ACM Press, 2003
- [22] Tolksdorf R., Glaubitz D., “XMLSpaces for coordination in web-based systems”, *proceedings of 10th IEEE Intl Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, IEEE Computer Society Press, 2001
- [23] Tolksdorf, R., “Coordination Technology for Workflows on the Web: Workspaces”, *proceedings of 4th International Conference on Coordination Models and Languages (Coordination2000)*, Springer LNCS, pp 36–50, 2000
- [24] RosettaNet, “RosettaNet home page”, <http://www.rosettanet.org/RosettaNet/Rooms/DisplayPages/LayoutInitial> , last accessed October 2004
- [25] SAP, “SAP Solutions web index page”, <http://www.SAP.com/solutions> , last accessed October 2004
- [26] Workflow Management Coalition, “Workflow Standard – Interoperability: Wf-XML Binding”, *WfMC Document Number WfMC-TC-1023*, 2000
<http://www.wfmc.org/standards/docs/Wf-XML-1.0.pdf> , last accessed October 2004
- [27] Object Management Group (OMG), “UML 2.0 OCL (Object Constraint Language) Specification”, 2003, <http://www.omg.org/docs/ptc/03-10-14.pdf> , last accessed October 2004
- [28] Ferguson, R., “Great expectations for SAP's NetWeaver”, *EWeek Magazine*, April 19, 2004, <http://www.eweek.com/article2/0,1759,1569238,00.asp> last accessed October 2004
- [29] World Wide Web Consortium (W3C), “XSL Transformations (XSLT) Version 1.0”, 1999
<http://www.w3.org/TR/xslt> , last accessed October 2004.

Visionary Management: Footstone of Sustained Firm Growth

Libing Shu¹, Shenghua Jia¹, Lei Peng²

¹ School of Management, Zhejiang University, Hangzhou 310027, China

² Institute of World Economics and Politics, CASS, Beijing 100102, China
Shualbert@163.com, jsh@zju.edu.cn, Penglei2005@163.com

ABSTRACT

As an important discipline of the learning organization, vision provides guidance about what core to preserve and what future to stimulate progress toward. It had been proved by the experiences of foreign companies' successful practice that visionary management was an actually new managerial philosophy and tool after the corporation reengineering theory to support sustained firm growth. It should be worthy our stated-owned enterprises and short-life's to learn and practice.

Keywords: vision, visionary management, performance, sustained firm growth

1. INTRODUCTION

Learning organization has been praised as the "Bible of management" and successful mode of future enterprise since Peter Senge (1990) put forward the learning organization theory. The origin of leaning organization theory is the outcome of deep thought for the sustained development in the dynamic and complex environment. As an important discipline of learning organization, vision provides guidance about what core to preserve and what future to stimulate progress toward. Visionary management has been a new organizational mind and management tool after the theory of corporation reengineering. Visionary management companies often achieve superior long-term performance and sustained growth, such as Hewlett-Packard, 3M, Johnson & Johnson, Procter & Gamble, Sony etc al. Hewlett-Packard employee have long known that radical change in operating practices, culture norms, and business strategy does not mean losing the spirit of HP way-the company's core principles. Johnson & Johnson continually questions its structure and revamps its processes while preserving the ideals embodied in its credo.

2. VISION AND ITS ORGANIZATION FUNCTION

2.1 What's the Vision?

Leaning and innovation is the best way to assure organization existence in dynamic and complex environment. Each organization should have five disciplines including building shared vision in order to create learning organization. Vision has become one of the most overused words conjuring up different images for different people. Vision is an image of the desired future. Collins and Porras (1996) argued that a well-conceived vision consisted of two major components: core ideology and envisioned future. Core ideology defines the enduring character of an organization - a consistent identity that transcends product or market life cycles, technological breakthroughs, management fads, and individual leaders. In fact, the most lasting and significant contribution of

those who build visionary companies is the core ideology. It likes the glue that holds an organization together as it grows, decentralizes, diversities, and expands globally. Any effective vision must embody the core ideology of an organization, which in turn includes two distinct parts: core values, a system of guiding principles; and core purpose, the organization's most fundamental reason for existence.

The core values stem not from market requirements but from the founder's inner belief. The key is not what core values an organization has but that it has core values at all. So a company should not change its core values in response to market changes but change markets if necessary to remain true to its core values. Core purpose is the organization's reason for being. An effective purpose reflects people's motivations for doing the company's work. It captures the soul of the organization, and doesn't just describe the organization's output or target customers. It is like a guiding star on the horizon - forever pursued but never reached. The role of core ideology is to guide and inspire, not to differentiate. Two companies can have the same core values or purpose. Core ideology needs to be meaningful and inspirational only to people inside the organization; it need not be exciting to outsiders. A clear ideology attracts to the company people whose personal values are compatible with the company's core values. Conversely, it repels those whose personal values are incompatible.

The envisioned future is what we aspire to become, to achieve, to create something that will require significant change and progress to attain. It consists of two parts: an audacious goal plus vivid descriptions. Visionary companies often use BHAGs (Big, Hairy, and Audacious Goals) as a powerful way to stimulate progress. All companies have goals. But there is a difference between merely having a goal and becoming committed to a BHAG - such as climbing Mount Everest. In addition to vision - level BHAGs, an envisioned future needs a vivid description of what it will be like to achieve your goals.

2.2 Organization Function of Vision

Vision has strong organizational function. Firstly, shared vision could strength organizational cohesion. Vision holds together all organization members who have different background. And it would change the relationship between the employee and organization. Employee's attitude to organization would change radically in a visionary company. They would think that firm isn't "their firm" but "our firm". Resentment to reform and passive obedience to vision would be replaced by active commitment and dedication to firm. Secondly, vision points out right direction for organization just like rudder of a ship, especially when organization encounters disorder or resistance in process of growth. Lastly, vision could inspire endless creativity and enterprising spirit of all members, and change the learning mode from adaptability learning to creativity learning. The essence of innovation is to reshape the world according to a special vision (Ikujiro Nonaka, 1991).

3. A VISIONARY MANAGEMENT MODEL

Visionary management is a key issue for all organizations. It's both a kind of managerial philosophy and a set of guidelines to form the basis for a never-ending journey toward competitive advantage. The visionary management process comprises of a cycle of successive phases, which can be distinguished into personal vision development, organizational vision development, situation analysis, strategy formulation, and planning and implementation (Rampersad, 1997). Figure 1 shows a model representation of visionary management. It begins with the formulation of personal vision, then expanding this vision to develop a common vision and core values of the whole organization. We should balance the personal vision and the organizational vision. The development of the personal and organizational vision does not only take place at the management level, but also at all lower organizational level. Everyone in the organization needs to be actively involved in this top-down and bottom-up visionary management process (Senge, 1990).

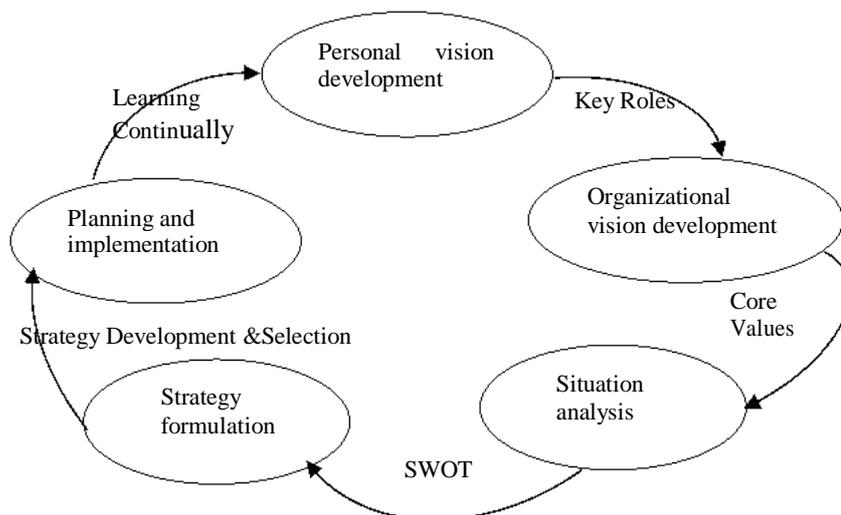


Figure 1 A visionary management model

3.1 Personal Vision Development

Each employee has to formulate his/her own personal vision. Personal vision development acts as key roles in the process of organization vision formation. Personal vision statement is a description of where are you going, which values and principles guide you to reach that point, what you want to help realize in your life, what ideal characteristic you would like to have. The key roles regard the way you fulfill or want to fulfill several roles in your life to realize personal mission. According to Covey (1993), personal vision helps to discover your inner feeling and clarifies what is important to you. You can eventually adapt vision in case your living conditions or thinking patterns change in the course of several years.

3.2 Organizational Vision Development

An effective developed organizational vision gives direction to personal ambitious and creativity, creates a climate through drastic changes, strengthens belief in the future, and releases energy in people. Organizational vision forms an important management instrument that expresses the soul of the firm. It gives direction to firm growth and functions as a compass and road map. The convincing propagation of an active, inspiring, challenging, and fascinating vision usually lead to more effort, satisfaction, and commitment. An organization will be successful if it succeeds to create a sense, a meaning that releases energy in people, raises involvement and puts people in movement. The organizational vision is linked to core values in order to strengthen the one-mindedness of the employees. The core values determine what approach is used to realize the vision. After all, core values in an organization

usually inspire commitment, loyalty, and devotion in all parts of the organization.

3.3 Situation Analysis

Situation analysis is the following step in the visionary management process and regards the evaluation of the strategic position of the organization and document of the organization's objectives. From strategic perspective to analyze the strengths and weaknesses of the organization (internal) and the opportunities and threats from the environment (external) is very useful to strategy formulation. Benchmarking is an important instrument here that is focusing on improving performance with respect to the leading competitor.

3.4 Strategy Formation

Alternative strategies are developed and choices are made, which based on the visible gap between the present and desired situation. Strategy indicates how the objectives can be realized and which choices should be made. It is the answer of SWOT analysis. Strategy should be translated into critical success factors (CSFs) to carry out it effectively. These key factors determine the competitive advantage of the organization because they are strongly related to its core competence. A CSF is that in which the organization must be outstanding to be able to survive, or that which is of decisive importance to the success of the organization. Examples of CSFs are: well-motivated and skilled employees, customer orientation, time to market, efficient dealer's organization, and good customer service etc al. They should be determined through brainstorming with the entire management team.

3.5 Planning and Implementation

In the last phase, the chosen strategies are thus translated into operational plans for the different departments or units within the organization, after which they will be implemented. The actions to be undertaken are worked out in the operational plans for the different business function. The use of project management techniques is usually necessary during the process, as well as creating awareness for change. Continually learning and communication will help to form new personal vision and to be harmony with organization vision.

After this phase the visionary management will start again. Visionary management is not a one-time activity, but a process of continuous improvement. By making visionary management as a part of your daily routine, you will integrate it into all aspects of your work.

4. VISIONARY MANAGEMENT AS THE SUPPORT OF FIRM GROWTH

Visionary management provides ideology and guidance for firm growth, just like the lighthouse of navigation.

Firm performance provides finance guarantee and possibility of growth, just like catalyzer of chemical reaction. We construct a relational model of visionary management and sustained firm growth. In figure 2, visionary management and firm performance act on firm growth through path I and II respectively. There are a strong correlation between visionary management and firm performance (broken line in figure 2), and which had been confirmed by many researches. So visionary management supports sustained firm growth on one hand through path I, on the other hand through path II in virtue of mediating variable performance. Consequently, we hold that visionary management is the footstone of sustained firm growth on the earth. Vision should be matched with firm growth. And the firm growth is changing continually along with internal and external environment, so visionary management is also a dynamically adaptable process.

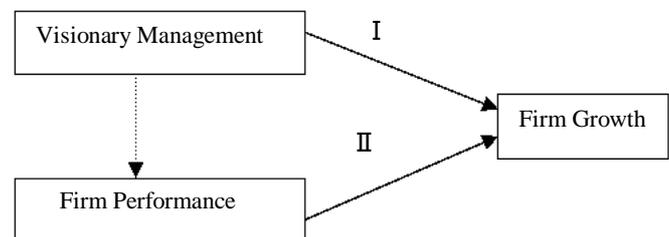


Figure 2 Visionary management and firm growth

Many world companies were pride of having definite and vivid vision, for example Microsoft, Hewlett-Packard, and Sony etc al. Effective management is the radical reason of sustained firm growth. Scholars also proved the relationship between visionary management and sustained firm growth from different aspects. Peter Senge (1990) pointed out that vision was an important factor to steer organization growth. The learning organization couldn't exist without vision. John P. Kotter (1995) carefully observed over 100 companies with different size, nationality and phase by ten years. And he found that some companies' change and transformation were very successful, but others were actually opposite. So he indicated that vision served as a connecting link between the preceding and the following in the process of leader change. Lack of vision and effective visionary management would influence the firm growth and successful transition. A lot of experiences of transitional companies also made clear that transitional endeavor were usually to fail without a solid visionary plan. The conclusion of Cranfield Global Executive Competencies Survey demonstrated that it played an important role for firm growth by forming a meaningful vision and future direction. Collins and Porras (1994) studied companies in their book of "Successful Habits of Visionary Companies" and found that they had outperformed the general stock market by a factor of 12 since 1925. Collins and Porras (1996) observed many company's growth status spending six years and concluded that the match of core values and visionary plan was the source of company life force. These companies holding clear

vision often have higher learning speed and accordingly bring higher firm performance.

5. CONCLUSION

(1) Visionary management as another bran-new managerial philosophy and tool after cooperation reengineering theory, which had been widely used by big companies in developed countries. The concept was just introduced to China companies and rarely practiced. Using foreign successful visionary management experience for reference has positive meaning to promote our firm performance.

(2) Visionary management provides source of power and new management tool for firm growth. It also reins to go forward along right direction. The strength of maintaining status couldn't be broken to buck for firm growth if there is no vision as road map to lead firm development.

(3) Visionary management is especially more meaningful to two kinds companies that are the state-owned companies and short-life companies. Now our state-owned companies are standing at the crossroad of reform and transition. Oversea experience has proved that effective visionary management could advance the probability of successful reform and transition. The relationship between visionary management and

sustained firm growth should cause thinking by large quantity of short-life companies.

REFERENCES

- [1] D. Keith Denton, "Building a shared vision", *Advanced Management Journal*, Vol. 62, No. 1, pp35-40, 1997.
- [2] Gray Hamel, C. K. Prahalad, "Competing for the future", *Harvard Business Review*, Vol. 72, No.4, pp122-128, 1994.
- [3] Hubert K. Rampersad, "A visionary management model", *The TQM Magazine*, Vol.13, No. 4, pp211-225, 2001.
- [4] Ikujiro Nonaka, "The knowledge-creating company", *Harvard Business Review*, Vol. 69, No. 6, pp96-104, 1991.
- [5] James C. Collins, Jerry I. Porras, "Building your company's vision", *Harvard Business Review*, Vol. 74, No. 5, pp65-77, 1996.
- [6] John P. Kotter, "Leading change: why transformation efforts fail", *Harvard Business Review*, Vol. 73, No. 2, pp59-68, 1995.
- [7] Rigby Darrell K., "What's today's special at the consultants' café", *Fortune*, Vol. 138, No. 5, pp86-87, 1998.
- [8] Senge P. M., *The Fifth Discipline: The Art and Practice of Learning Organization*, New York: Doubleday, 1990.

A Risk Identification Method of Virtual Enterprise

Min Huang¹, Fei Xu¹, Xing-wei Wang¹, W.H.Ip², K.L.Yung²

¹ Box 135, Faculty of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, China, mhuang@mail.neu.edu.cn

² Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, hongkong.hk.polyu.edu.cn

ABSTRACT

Virtual enterprise is the potential mode of future enterprise under global economic environment and the risk management for it is a popular research area recently. Risk identification is the key phrase of risk management. Considering the risk identification problem in the virtual enterprise, a new identification method based on the BP network is provided in this paper. Using this method and the data related to the problem, the possible to take place of the risks could be obtained by training the neural network, so that risks can be investigated in time and managed effectively. The simulation results prove its validity.

Keywords: virtual enterprise, risk management, risk identification, BP network learning algorithm

1. INTRODUCTION

Global economic environment is experiencing tremendous changes in recent decades as information technology and its application advance rapidly. Market demand is changing even rapidly and becoming more difficult to predict as the global market competition heats up. Under such background Virtual Enterprise (VE), a new organizational mode for enterprises, is gradually replacing traditional modes of enterprises [1] for it's more adaptive to new challenges. VE enables enterprises related to one product to combine together to form a temporary economic body so that they can grasp new opportunities and response to latest market demand as soon as possible. The new mode is considered profitable though it also brings in risks and uncertain factors, which might have harmful influences on enterprises involved. Traditional risk identification methods, depending on personal experiences and subjective judgments, are mostly qualitative [2,3]. They do not perform well in identifying new risks under the recent economic environment. Such shortage calls for a better risk identification method to ensure the secure operation of enterprises. Data mining (or knowledge discovering) technology, a process based on artificial intelligence and database technology in which some models or modes that are useful to data owner(s) are extracted from a large amount of data, is providing a new thought for resolving this problem. The present functions of the technology mainly include mode identification, association analysis, sorting, predicting and cluster analysis, etc.[4, 5] on the basis of algorithms and technologies of artificial intelligence. In the research risk identification problem is solved by data mining method based on BP network and with the help of Statistical Analysis System (SAS) software, the application is proved to be satisfactory.

2. PROBLEM DESCRIPTION

Successful risk management depends on scientific

procedures. The process of risk management includes risk identification, risk estimation, risk evaluation, risk programming and effective risk control based on the previous steps. During risk identification, the primary step of risk management, it is necessary to determine the general objective of enterprise and the partial objectives needed for realizing the general one. Based on this, risks factors within the general objective and partial objectives are investigated and analyzed, which can be achieved through survey/ questionnaire, organization chart analysis, financial statement analysis, checking chart analysis, process analysis, input-output analysis, sensitivity analysis, accident-tree analysis, hierarchical holographic modeling analysis and expert investigation, etc [2,3]. These methods:

- 1) Require investigator or analyst's deep understanding of enterprises and long-term information accumulation;
- 2) Require long-term investigation and analysis;
- 3) Respond slowly to environmental changes.

However, the VE is formed dynamically, which means operators can only acquire information and data of pre-existing VE and original enterprises. Consequently, VE should decide based on the data whether the current risks are possible to take place. A new method towards such a problem is introduced in the research --- BP-network-based risk identification of VE. According to this method, every possible kind of internal or external risks is firstly analyzed and the information concerned defined; then the possibility of each kind of risk is got according to data of pre-existing VE, original enterprises and information of current VE.

3. NEURAL-NETWORK MODEL

BP-network is one of the most widely-used neural

networks at present. Before putting BP-network into practice, the number of hidden hierarchies and that of units in each hierarchy have to be fixed so as to determine a network topology. Figure 1 shows a 3-hierarchy forward neural network with only one output nodal point. From left to right aligns input, hidden and output hierarchy. Combination function between hierarchies is a lineal function. The net input of each unit j in hidden or output hierarchy is:

$$I_j = \sum_i w_{ij} o_i + q_j \tag{1}$$

Among the function, w_{ij} is the linking weight from previous unit i to unit j ; o_i represents the output of unit i ; q_j denotes the offset of unit j , which also presents a threshold here. Appropriate offset help to change the activity of the neuron.

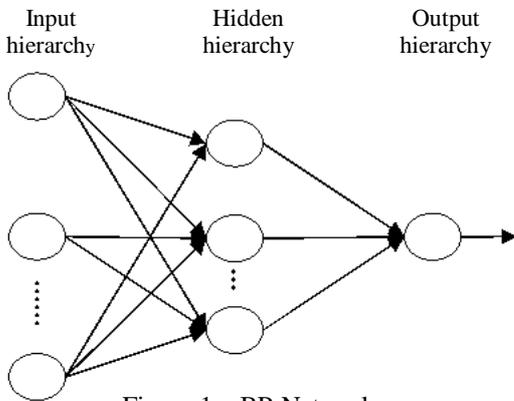


Figure 1 BP-Network

After estimating the net input in each unit, an activating function has to be chosen to estimate the output value. Simoid activating function, the most common type, is chosen here:

$$o_j = \frac{1}{1 + e^{-I_j}} \tag{2}$$

In BP-network training, classical BP algorithm is applied, showing a forward transmission of input and a backward transmission of error. Linking weight and offset are modified by calculating the error between the neural network predicted and the value of sample. If linking weight and offset end up in convergence after a number of iterations, the learning process stops; otherwise re-training or adjustment of the network structure is needed.

4. APPLICATION AND KEY PROBLEMS

SAS, one of world’s most famous statistical software, is developed by SAS Institute of the US and used to help realize the process of risk identification in this research. As the technology of data mining thrives these years, SAS institute has developed for this field a tool called SAS/EM (Enterprise Miner) , which then became a best data mining software because of its integrated and

powerful statistical function. The outcome of data mining also shows a progress since the software has been worked on as a product and improved. Its user-friendly interface also brings in convenience to the data mining process [6-8].

Before data mining, should be processed for preparation so as to ensure validity and high accuracy:

1) Data Set Division

Over-fitting is considered a very significant problem in neural network training. It can be defined as a process in which a neural-network model fits the current data set too well(this might even mean a transitional fitting) to adapt to a new set flexibly. A common way to avoid over-fitting is to divide the data set into a training set and a valid set so that the training process may be applied to two sets at the same time. Training set is for BP-network training and valid set for evaluation to the model. Taking the average error as a performance measurement of training process, it shows a continuous decrease of average error in training set as the training process goes on while the average error in valid set decreases first and then increases. This can be explained as a consequence of over-fitting of neural-network to training set. In this research, as a principle, the weight value, when the average error of valid set is the lowest, is selected as the final solution.

2) Data Standardization

Usually statistics within a data set is not acquired under a same commensurability but by individual method and unit of measurement, which indicates that different input variables of BP-network may have unbalanced influences on target variables. The unbalancing situation among variables caused by disunified measurement is obviously irrational and it might waken some variables of great significance. In order to avoid any mistake caused, the following method of standard deviation (each variable is divided by the standard deviation of sample) is applied. The standard deviation of variable number k that is denoted as variable X_k is estimated in the following formula:

$$s^{\wedge}_k = \left(\frac{1}{n} \sum_{i=1}^n (x_k(i) - m_k)^2 \right)^{\frac{1}{2}} \tag{3}$$

m_k , denoting the average of X_k , is estimated through sample average

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_k(i) \tag{4}$$

Ultimately the inaccuracy brought by disunified measurement is eliminated by calculating

$$x'_k = \frac{x_k}{s^{\wedge}_k} . \tag{5}$$

After the two steps above, BP-network training begins.

The training can be carried out by the Neural Network Function in SAS/EM.

5. SIMULATION ANALYSIS

Simulation data includes 5822 observation results gained per hour by per-existing enterprises, each of which contains 83 input variables and 1 target variable. 83 input variables are enterprise-related data, including the number of partner enterprises, communicative efficiency within partners, information-safety rate and product-purchasing rate, etc. The target variable denotes the possibility of the risk of product quality. The data set is divided randomly into a training set, which contains 70% (4075 pieces) of observation results and a valid set, which contains 30% (1747 pieces) of observation results. There are many methods for evaluating the model, such as error-classifying rate, average error, profits/loss, etc, but in regards to the nature of risks, wrong prediction of the possibility of a risk, and its deviant receive more attention. Consequently error-classifying rate and average error are chosen to be the evaluating measurement to BP network model.

BP network training is a repeating and continuously adjusting process, which needs several parameters to be

units in hidden hierarchy, learning rate and numbers of iteration. In these parameters, the initial value of weight almost have no effect on the training result. The numbers of iteration should be proper; too small the training will be not convergence, too big the training will last long time. However, the numbers of units in hidden hierarchy and learning rate effect the accuracy and process of training largely. Figure 2 and 3 show respectively the influence of them on average error in both training and valid set.

The numbers of units in hidden hierarchy is crucial to BP network training because it directly affects the quality of prediction. Insufficient units may result in divergence or dissatisfactory accuracy of prediction while too many units sharply increase training time and arouse more difficulties in learning process. As the training goes further, it is shown in Figure 2(a) that under the same parameters, the increasing number of units in hidden hierarchy results in a smaller model-predicted average error but too many units takes a much longer time for training. Also, the Figure 2(b) shows the numbers of units in hidden hierarchy should be proper. So, in general, 16 is chose as the numbers of units in hidden hierarchy when Figure 2 (a) and (b) is considered synthetically.

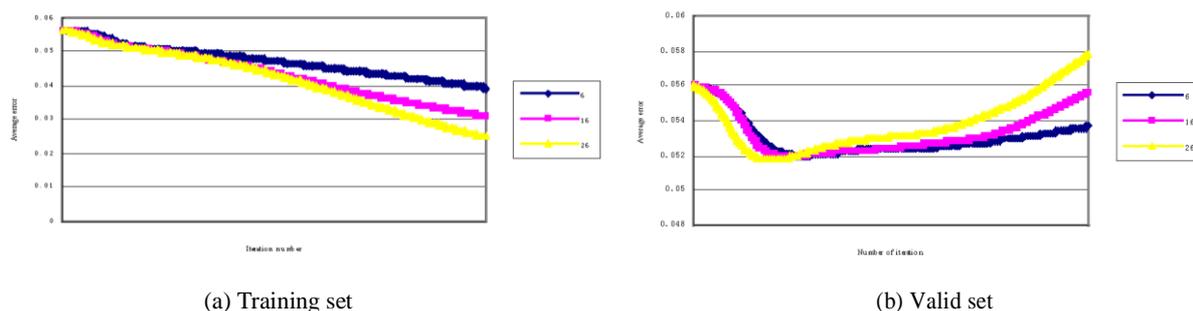


Figure 2 The influence of the numbers of units in hidden hierarchy on average error

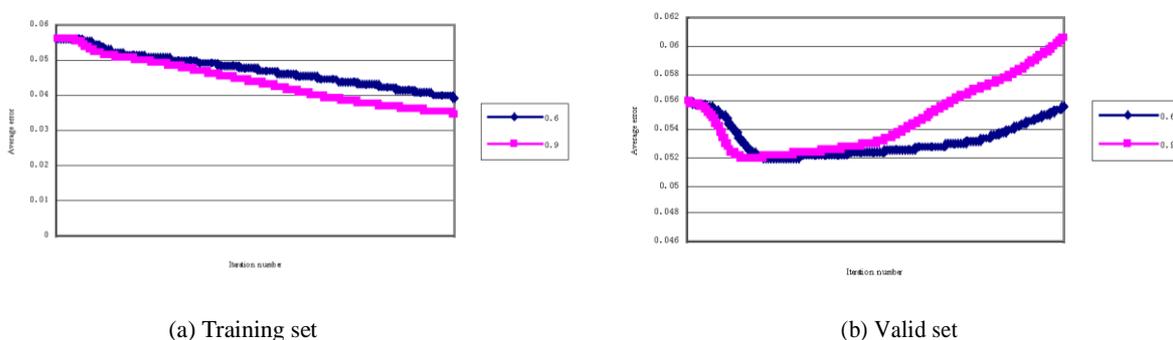


Figure 3 The influence of learning rate on average error

setting first: the initial value of weight, the number of

Figure 3 shows the influence of learning rate on the training process. Figure 3(a) shows that lower learning rate results in slower training but an excessive learning rate (bigger than 8) may make the training process swing among unsuitable solutions or even results in divergence. Figure 3(b) shows although the average error of the learning rate 0.6 is higher in training set, it is lower in valid set. Hence, 0.6 is chose for learning rate.

Figure 2(b) and 3(b) also show that the average error in valid set decrease first and then increase, which is the “Over-fitting”process. To avoid this, the weight of the smallest average error point in valid set is used as the final training result.

Figure 4 shows the training curve and valid curve

illustrating the changing average error in neural network training. The vertical line shows the smallest average error point in valid set, to which the weight corresponds is used as the ultimate result.

Based on the analysis above, the structure of BP-network model and parameters for training are determined as shown in Table 1.

Table 2 lists out the result of model evaluation. It shows that the average error and error-classifying rate are small enough in both training and valid set, which proves the model’s excellent ability in risk identification

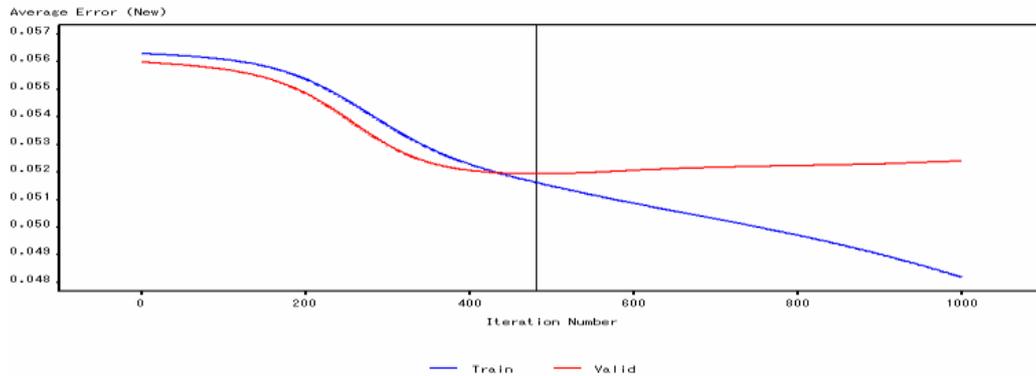


Figure 4 BP training process

Table 1. Structure of BP network model and training parameters

Number of input hierarchic units	83
Number of hidden hierarchic units	16
Number of output hierarchic units	1
Activating function of hidden and output hierarchy	$f(x) = \frac{1}{1 + e^{-x}}$
Combination function between	$\sum_i w_{ij} o_i + q_j$
Learning rate	0.6
Iteration number	1000

Table 2 Results of BP identification

Measurement Data set	Average error	Error-classifying rate
Training set	0.0516240742	0.0598773006
Valid set	0.0519343988	0.0595306239

6. CONCLUSIONS

It’s a brand-new attempt to solve risk identification problem by BP network. Simulation results have proved the method’s validity. Some excellent software for data mining also makes the application of this method possible. With an accurate identification of risks, enterprises are capable of taking more effective measures in their risk management.

ACKNOWLEDGEMENTS

The authors wish to thank the support of the National Natural Science Foundation of China (Project no. 70101006, 60473089, 60003006), Liaoning Province Natural Science Foundation of China (Project no. 20032019, 001015), the Scientific Research Foundation for the Returned Overseas Chinese Scholars by SEM of China, Modern Distance Education Engineering Project

by MoE of China, and the Research Grants Council of Hong Kong, China (Project no. PolyU 5167/99E).

REFERENCES

- [1] Huang, M., Yang H.M. and Wang X.W., "Genetic algorithm and fuzzy synthetic evaluation based risk programming for virtual enterprise", *Acta Automatica Sinica*, Vol. 30, No. 3, pp449-454, 2004.
- [2] Haimes, Y. Y., *Risk Modeling, Assessment, and Management, 2nd ed*, New York: Wiley, 2004.
- [3] Bier, V.M., Haimes, Y.Y., Lambert, J.H., Matalas, N.C. and Zimmerman, R., "Survey of approaches for assessing and managing the risk of extremes", *Risk Analysis*, Vol. 19, No. 1, pp 83-94, 1999.
- [4] Han, J.W. and Kamber, M., *Data Mining: Concepts and Techniques*, Beijing: China Machine Press, 2001.
- [5] Hand, D., Mannila, H. and Smyth, P., *Principles of Data Mining*, Cambridge, Mass., London: MIT Press, 2001.
- [6] SAS Institute Inc., "Business intelligence systems and data mining", *SAS Institute White Paper*, Cary, NC: SAS Institute Inc., 1997.
- [7] SAS Institute Inc., "Using data mining techniques for fraud detection", *SAS Institute Best Practice Paper in Conjunction with Federal Data Corporation*, Cary, NC: SAS Institute Inc., 1999.
- [8] SAS Institute Inc., "Finding the solution to data mining: a map of the features and components of SAS enterprise miner TM software version 3", *SAS Institute White Paper*, Cary, NC: SAS Institute Inc., 2000.

A VOC Based Approach to Support Virtual Organizations

Hua Zhou¹, Min Liu², Cheng Wu³

¹ Department of Automation, Tsinghua University, Beijing 100084, China
Zhouhua02@tsinghua.org.cn

² Department of Automation, Tsinghua University, Beijing 100084, China
lium@mail.tsinghua.edu.cn

³ Department of Automation, Tsinghua University, Beijing 100084, China
wuc@tsinghua.edu.cn

ABSTRACT

Employing IT as the key enable technology, virtual organization (VO) is primarily characterized as being a network of independent, geographically dispersed organizations (member organization) providing electronic services via Internet. To align these services effectively, one of the main challenges is to model cooperation in VO and provide correspondent management tools. Here we present a VOC (VO Structure-Organization Resource-Character) approach to model and run VO. VOC consists of three models. VO Structure model describes how VO functions in terms of VO Role (VR), Protocol; Organization Resource model describes potential service providers (potential member organization) capable participate VO. Character Model describes dynamic binding relationship between VO role and member organization. Following the VOC model, a platform supporting the design, administration, and running of VOC is given.

Keywords: Virtual Organization, VOC model, VO Structure Model, Organization Resource Model, Character Model

1. INTRODUCTION

A virtual organization (VO) is primarily characterized as being a network of independent, geographically dispersed organizations with a partial mission overlap^[1,2]. Each VO is built to carry out some processes and achieve certain goals, and as the products and services provided by a virtual organization are dependent on innovation and are strongly customer-based, the organization structure must be adaptively dynamic. Within the network, all partners provide their own core competencies as deliverable services, and in a networked computational environment, some of these services are delivered in electronic form, for example, web services. To make a VO work flexibly and smoothly, a proper VO model and correspondent supportive running environment that can align these e-services are definitely necessary.

In this paper, we present a VOC (VO Structure-Organization Resource-Character) based approach to model and run VO. The VOC approach consists of a VOC model and correspondent supportive platform.

In section 2, we illustrate VOC model and some core processes, in section 3 a VOC-based VO management platform prototype is introduced and several correspondent core processes is then illustrated. Then we present a collaborative manufacturing example in section 5. Finally we conclude some characteristics of the VOC approach.

2. VOC MODEL

In a VO, each member organization (MO) has to assume

certain duties, deliver prescriptive services, join and leave VO dynamically. In VOC approach, we model the whole VO in several different models with each one focusing on different aspect of VO.

VOC model consists of three sub models:

- ò VO Structure Model (VSM)
- ò Organization Resource Model (ORM)
- ò Character Model (CM)

The VOC meta model is described in Figure 1.

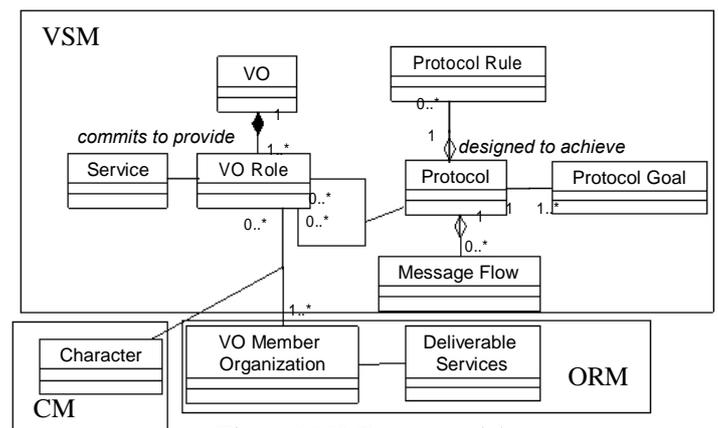


Figure 1 VOC meta model

2.1 VSM

The VO Structure model (VSM) describes VO structure in terms of VO Roles and Protocols. VO Role (VR) acts as a duty placeholder and commits to realize certain services in VO. VR will be dynamically assigned to MO

at VO design time and runtime (MO being the undertaker of VR). VR undertaker must realize the services VR commits to provide. VR cooperates with each other to achieve VO goals. The cooperation relationship between VR is modeled as Protocol. Protocol includes protocol rule and Message Flow. The former represents constraints that VR has to comply with. The latter illustrates message template and message sequence exchanged between protocol participants (VR). Protocol is instantiated as conversation which maintains the concrete context and threads participants will refer to frequently during cooperation.

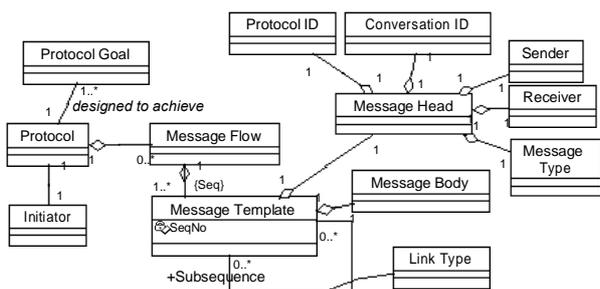


Figure 2 Message Flow

VR models static aspect of VO while Protocol models dynamic aspect. Each Protocol is designed to satisfy certain cooperation goals. For example, we can design a protocol based on Contract Net with its goal being task delegation. During runtime, VR undertaker can only choose from prescriptive protocols to cooperate with other VR, and follow exactly the Message Flow (Figure 2). Each protocol has an Initiator who is responsible for creating a protocol-based conversation. Each Message Template has a head and body. The Head includes message related protocol ID, conversation ID, sender VR, receiver VR and message type. The Body contains message content. Conversation participant must send message in according with the sequence and template specified in Message Flow to help each other indicate, understand, respond message quickly and properly.

2.2 ORM

Organization Resource Model (ORM) describes MO's capability and deliverable services. ORM of VO contains all potential MOs of a VO. Organization in ORM can be deputed VR dynamically by VO Administrator.

2.3 CM

Character Model (CM) is responsible for binding ORM and VSM. To make a VO functions as will, each VR has to be assigned to certain MO. We name this kind of 'Play' relationship between VR and MO as character. A MO can assume several characters at the same time.

3.VOC-BASED PLATFORM

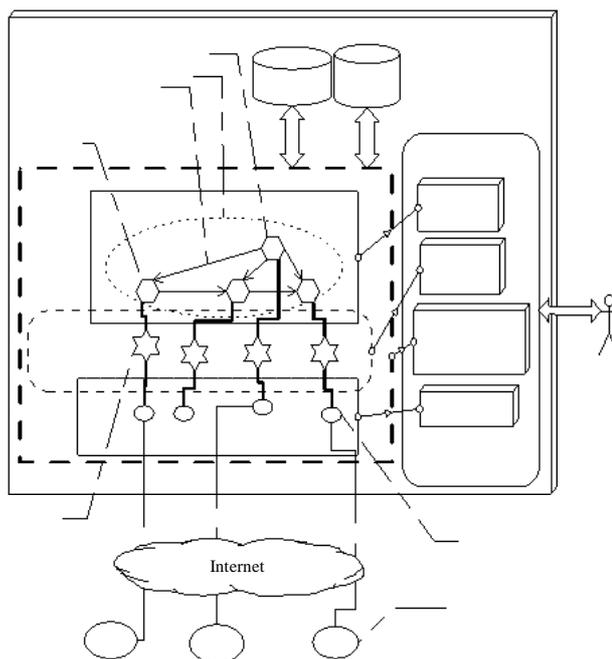


Figure 3 VOC based Platform structure

Based on VOC model, we developed a VOC based platform prototype (Figure 3). The core VOC model element is realized as Java object running in classified pools: VSM Pool; Character Object Pool; MO Agent Pool. Each kind of pool is managed by a correspondent software component.

By using the platform, we can design and run VO. The platform includes a VOC Model DB, a VO Instance DB, and VO Tools.

ò VOC Model DB

VOC Model DB stores all VOC model data, including VSM, ORM and CM. VOC model works just like a class in Object-Oriented technology. A running VO is an instance of a special VOC model and consists of different types of software objects (or components) living in correspondent object pools.

ò VO Instance DB

VO Instance Db stores all VO instance related data such as VO start time, VO original VOC model, VO MO etc.. By providing VO Instance DB, we can adjust VO in a case-based approach and thus providing a flexible way of managing VO.

ò VSM Mgr

VSM Mgr is responsible for VSM management, including VSM creating, saving and dynamic adjustment, VSM Object initiation and life cycle management. VSM Mgr is the manager of VSM pool. There are four types of object in VSM Pool:

V VO Object

A VO object is the controller of a running VO. When a VO is initiated the system will create a VO object and this VO object will be responsible for managing the whole life cycle of VO and create, manage, destroy other VSM object in this VO as needed.

V VR Object

Each VR in VSM will create a VR Object at VO runtime.

V Protocol Object & Conversation Object

Each protocol in VSM will create a Protocol Object at VO runtime. When a protocol-based interaction is initiated, a Conversation Object is created to manage it. Conversation Object acts as a communication channel between VO Object which can only send messages through the interface Conversation Object provided and in the protocol specified form and sequence. Conversation Object can get message and rule related information from Protocol Object at runtime. For there may be tens of thousands conversation being running at the same time, saving protocol related information in Protocol Object is wise choice to avoid frequent DB access.

ò ORM Mgr

ORM Mgr manages ORM, and can instantiate MO Agent Object (MAO) according to the registered MO in ORM. Any organization that wants to participate in the VO must be registered in ORM. There are two ways of registering MO: active way and passive way. In active way, organizations must register themselves in ORM by providing their own capacities and deliverable services information. In passive way, VO Administrator has to search above information and identify appropriate organizations, contracts with them and registers them manually in ORM. Each registered organization will get a reference (though Web Service) to a MAO which is created by ORM Mgr in MO Agent Pool. When a registered organization exits from ORM, the correspondent MAO will terminated.

V MO Agent Pool

A MAO can only lives in MO Agent Pool. Each MAO is just a proxy of registered MO in VOC Platform and is exposed to related MO as an accessible web services. MAO provides an effective way of communicating with MO and is independent from the internal realization structure of original MO. We also provide an interface for MAO so that it can interact with Character Object smoothly.

ò Character Binder

Character Binder manages the binding between MO and VR. A VO can not be instantiated until each VR has been assigned to a concrete MO. A MO-VR pair is represented as a character and instantiated as a Character Object (CO). Thus, within a VO, MO can cooperate with each other in this way: MOÆMAOÆCOÆVR ObjectÆConversation

ObjectÆMAOÆMO. This indirect way of interaction may bring some efficiency problems but illustrate a flexible way of reconfiguration. By making all objects active in pool (RAM), we can alleviate the efficiency problem to some extent.

ò VO Admin Center

VO Admin Center is the center controller of the whole VOC-based platform. VO Administrator uses it to manage all VO running in the VOC-based Platform. VO Admin Center calls VSM Mgr, Character Binder, ORM Mgr to create, destroy, and adjust VOC model and instantiated VO as well.

3. VO CORE PROCESS

To build VO based on the platform in an effective way, we identified some core processes in constructing VOC model and running VO. These are processes the platform supports and each MO has to comply with. We illustrate several processes below:

ò MO Registering Process

MO can participate in special VO by first registering itself in VOC-based Platform and then get a MAO reference.

ò MO Terminating Process

MO can only terminate from running VO after issuing MO-TERMINATION message to all other VRs and complete works on hand. Thus other VRs won't send it any more messages after their completing to-be-terminated MO related cooperation work. The platform will then destroy related CO, MAO.

ò MO Entering VO Process

After a special MO is selected to assume a VR, it has to realize all services specified in VR specification. First, the platform will send the target MO a VR specification identifying all services and protocols MO has to realize in detail. Then MO must design and realize it, and then answer the platform with a SERVICE-READY message. Only after this process, the platform can confirm the VR-INSTANTIATION-READY status of VR.

ò VSM Instantiating and Terminating Process

The platform first checks all its VR to confirm all of them being in VR-INSTANTIATION_READY status, and then creates the whole VO instance in this sequence: VO ObjectÆVR ObjectÆCharacter Object. To terminate a exist VO instance, the platform first check all VR to confirm all MO has issuing MO-TERMINATION message and then destroy the whole VO instance in this sequence: Character ObjectÆVR ObjectÆVO Object.

4. SAMPLE APPLICATIONS

By referring a production process in an aeronautic assembly factory, we construct our sample application

to verify the effectiveness of the platform (See Figure 4). The product experience 2 stages before assembled. Due to the cost-reduction problem, the first two stages is outsourcing to two different factories. But the contractual manufacturer is not fixed to allow more flexibility to purchase lower cost and higher quality. And production stages will change from time to time due to technology development. That is to say, the production routine and contractor list is not static. Thus the factory has to issue production task and cooperate with contractual manufacturers in an unstable production net. A production-net-adaptive platform supporting collaborative production task allocation is helpful.

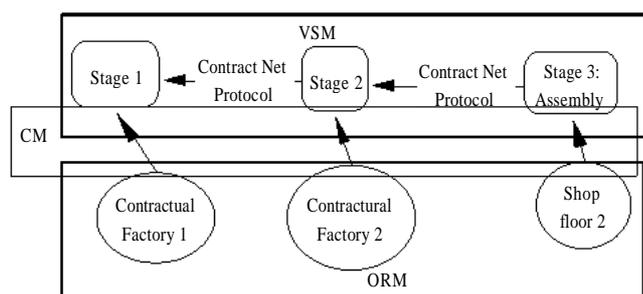


Figure 4 Collaborative Manufacturing Example

By applying VOC approach in this example, we can build Stage as VSM VR, and manufacturing contractor as MO, and the contractual manufacturing relationship as Character, thus we can adjust the collaborative relationship flexibly. Meantime, by designing proper VR Protocol suitable for production task allocation, we can change task allocation procedure as well.

5. CONCLUSIONS

The VOC based approach of managing VO has the characteristics as follows:

V Reusable

We divide the whole VO into three models with each one focusing on different aspects of VO. VSM models the static structure and cooperation relationship between MO. ORM works as a resource pool while VO Administrator can choose from. Character Model connects VSM and ORM to form a concrete VO. Thus we can conclude the reusable issue into these three levels. In VSM level, a historical VO structure model can be a reusable unit. For example, we can elicit a proved effective VSM into enterprise knowledge base and turn it into the original reference model of an upcoming VO (like the case of section 5). In ORM level,

a MO is shared between different VO and can be assigned different VR from different VO at the same time. A MO's Character history can identify what kind of VR (protocols, services) it is most suitable for.

V Reconfigurable

As MO cooperate with each other in an indirect way and the existence of VO Instance DB, VO Administrator can reconfigure the VO structure by adjusting VSM, ORM or CM. For example, you can change the cooperation constraints by applying different protocol between VR.

V Model driven

Each VO is initiated and run based on the VOC model. You can influence the VO behavior by just adjusting VO case model stored in VO Instance DB.

V Scalable

ORM works as a resource pool where all potential VO participants can enter and exit flexibly.

Future research will focus on dynamic VR delegation and dynamic VR selection.

ACKNOWLEDGEMENT

ICEB2004 is supported by the National Natural Science Foundation of China.

REFERENCES

- [1] Yates, J., Orlikowski, W.J., Woerner, S.L., "Virtual organizing: using threads to coordinate distributed work", *Proceedings of the 36th Annual Hawaii International Conference*, pp271-280, Hawaii, 6-9 Jan. 2003
- [2] Troy J. Strader, Fu-Ren Lin b, Michael J. Shaw C, "Information infrastructure for electronic virtual organization management", *Decision Support Systems*, Vol.23, pp75-94, 1998
- [3] IGLESIAS C. A, GARIJO M, and GONZALEZ J. C., "A survey of agent-oriented methodologies", *Proceedings of the 5th International Workshop on Intelligent Agents*, pp317-330, Heidelberg, Germany: Springer-Verlag, 1999.
- [4] WOOLDRIGE M et al., "The Gaia Methodology for Agent-Oriented Analysis and Design", *Journal of Autonomous Agents and Multi-Agent Systems*, Vol.3, No. 3, pp 285-312, 2000
- [5] FERBER J, GUTKNECHT O., "A Meta-Model for the Analysis and Design of Organizations in Multi-Agent Systems", www.madkit.org/publication

Refinement for Ontology Evolution in Virtual Enterprises

Li Li, Baolin Wu, Yun Yang

CICEC Centre for Internet Computing and E-Commerce Faculty of ICT, Swinburne University of Technology
PO Box 218, Hawthorn, Melbourne, Australia 3122
{lli,bwu,yyang}@it.swin.edu.au

ABSTRACT

Virtual enterprise is based on the premise that work should be done where it can be done most optimally. In virtual enterprises, geographical boundaries merge seamlessly. It enables organisations to act in a way of flexibility and ability to adapt to rapid changes on the fly. However, different parties in a virtual enterprise must understand each other before they go further details in business. Ontologies are such kinds of ideal baselines to assist parties to communicate. One of the essential research issues with ontology is how to deal with changes during their evolving cycle. Therefore, ontology refinement is a crucial component in ontology evolution. This paper presents a taxonomy structure focusing on the is-a relations. In particular, the concept of closeness measurement is introduced based on the “distance” estimation. An extended cluster analysis process is provided. According to the algorithm presented, a new concept is generated according to its attributes. Additionally, the refinement mechanisms for primitive operations are proposed. Unlike some other ontology refinement mechanisms which leave ontology consistency checking to human users after modification, our method emphasises the importance of consistency checking by applying description logics which is demonstrated based on the proposed ontology.

Keywords: ontology evolution, virtual enterprises, clustering, description logic, consistency checking

1. INTRODUCTION

An ontology can be defined as *an explicit specification of conceptualisation* [6]. Currently, the use of ontology for different purposes such as Web applications, Web semantics, information systems, are intensively addressed. The ontology is important because it is the foundation for different organisations to understand each other within or across enterprises. Feasible ontologies in virtual enterprises (VEs) are especially important because of VE’s characteristics: VEs only exist for a short life span as a temporary network of geographically distributed partners that cooperatively work together to share skills, costs, profits, risks, and markets, and, at the same time, decrease the investment [3]. VEs are not rigid organisational structures within rigid frameworks, but rather, heterogeneous ensembles, continuously evolving over time [13]. For this reason, a complete ontology is essential for VE formation because of its pressing time requirement and heterogeneity. However, there indeed exist cases when some concepts/attributes and their relations are out of date or cannot be accessed through the current ontology. According to Stojanovic [14], there are seven different requirements demanding ontology evolution. In addition, a six-phase evolution process is addressed. Other researchers [12,14] also note that ontology development is a dynamic process starting with an initial rough ontology, which is revised, refined and filled with details on an ongoing basis. That is to say, ontology refinement is important and needs more efforts throughout its entire lifecycle. In VEs where different parties might work on an ontology collaboratively, dynamically refining the ontology is likely to be more important than that in any other places.

Intelligent software agents (agents hereafter) are a suitable means of representing the partners of a VE [5]. Agents act on behalf of the enterprises or organisations that collaborate among themselves to achieve a specific goal under assumptions that they are goal-oriented, commitment-based and able to share skills, cost, profits, risks and markets [12]. We assume agents who are likely to form a VE have similar but slightly differing ontologies rather than completely different ones. This is reasonable in the real world for different partners having something in common within the domain.

Some work has been performed on tackling ontology refinement. Noy’s work [10] talks about that both *merging* and *alignment* are two processes which are usually used to handle the ontology refinement. These two processes are illustrated in PROMPT [11], a semi-automatic approach to ontology merging and alignment where users’ intervention is required. In his book [8], Maedche mentions an essential part of ontology engineering ontology learning. The development of the taxonomic backbone of the ontology is also involved. The clustering analysis mainly discussed in natural language processing previously, has been highlighted from learning the taxonomic relation perspective.

Generally speaking, existing refining proposals fall into two categories. One is to investigate ontology dynamic changes and the corresponding management from the knowledge engineering perspective [1,4,8,14]. The other is from the context of plan execution [9]. However, in light of their proposed approaches, a new term’s relationships with existing concepts are specified/semi-specified by human users. By doing so they avoid indicating how they get to know the relations between con-

cepts. That is to say that they fail to address how they get to know the relations between new terms and the existing ones and whether the ontology is still consistent after some modifications.

To comply with the requirements, our main task will be on adjusting the proposed ontology structure to any changes of concepts and their relations while at the same time processing consistency checking with the source ontology. We use a statistics-based approach, especially the cluster analysis method first to obtain a corresponding classification of the terms (in the is-a concept hierarchy). Then we use description logics (DLs) [2,7] to check if the newly added concepts are consistent with the original ontology structure by considering their satisfiability and subsumption. The novel contribution of this paper is to provide a method that operates in a resilient manner in ontology refinement, and at meantime, consistency checking complying with the ontology evolution life cycle to guarantee that any application has a sound knowledge of the world.

We do not distinguish terms between parties, partners, participants and agents in this paper only if they would convey much more meaning in certain circumstance.

The rest of this paper is organised as follows. The next section first presents a taxonomy structure and then addresses the mapping between concept relations and distances. Section 3 proposes refinement mechanisms for the add and delete primitive operations. Section 4 discusses consistency checking of an ontology by using DLs. Section 5 illustrates our approach with an example. Finally, section 6 concludes our work.

2. CLUSTER ANALYSIS

Term *cluster analysis* [15] actually encompasses a number of different classification algorithms. Two things are worth noting here. One general question facing *cluster analysis* in many areas is how to *organise* observed data into meaningful structures— taxonomies. Another is a suitable algorithm for a specific question. In this paper, concepts are organised as the higher the level of concept aggregation the more abstract these concepts are in the respective class. We borrow term “distance” to show close/loose relations in the paper. In addition, we use the term “new cluster” to notate a new generated concept if needed in a dendrogram. It will be discussed in more detail in section 3.

2.1 Taxonomy Structure

A concept definition, along with its relations and others (if they exist), is more likely to be described in a taxonomy (the “is-a” hierarchy) structure. The most central relations are the is-a relation (concept— concept) and property-of relation (concept— attribute). Like in Figure 1, nodes in squares with capital letters such as *A, B, C, D, E, and F* are abstract concepts whereas other nodes

notated in circle with lowercases are concrete ones. The difference between an abstract concept and a concrete one lies in that the abstract one subsumes sub concepts. In this paper, concrete concepts are specified with primitive or common knowledge in terms of attributes and properties. For instance, in a general “transport” ontology which describes different transports such as “ship”, “airplane”, “bus”, and “train” etc. and their relationships, concepts like “ship”, “airplane”, “bus”, and “train” are abstract concepts, while “mph”, and “capacity” are concrete concepts to define the attributes of a specific transport. Clearly, concept *B* subsumes both *E* and *F*, whereas *B* is subsumed by another concept *A*. And lowercase *f* is a property-of concept *A*. Obviously, the higher the concept is, the more abstract it is. Ontology structure in this paper is described in the “is-a” hierarchical structure. And we will refer to Figure 1 again in subsequent sections. Understanding how a dendrogram is constructed, and how it should be interpreted, is helpful to understand ontology refinement by using clustering analysis. The next subsection will discuss Figure 1 in-depth from the cluster analysis perspective.

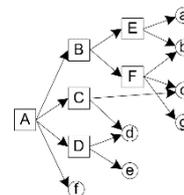


Fig. 1. Taxonomy structure of concepts

2.2 Mapping Concept Relations and Distances

Every concept, no matter it is a concrete or an abstract concept, has a set of attributes with different levels of granularity. Generally speaking, a concept can be expressed by a set of triples of (attribute, value, weight), namely $A_{att} = \langle (a_1, v_1, w_1), (a_2, v_2, w_2), \dots, (a_m, v_m, w_m) \rangle$, where a_j is an attribute and v_j is the corresponding value. Here w_j ($j \in [1, m]$) is represented by a real number in range $[0, 1]$ which is an estimation standing for to what extent an attribute getting a specific value (1 means matching perfectly, while others mean they can be approximately equal to those values). An instance such as (shape, circle, 0.8) means that a concept has attribute “shape” with value “circle”, but not a 100 percent circle, the closeness estimation is only 0.8 instead of 1.

In order to reduce redundancy, a value matrix (*VM* in short), a corresponding attribute vector (*AV*), and its corresponding weight matrix (*WM*, in which 0 means corresponding attribute is *N/A*) are introduced below. It is easy to get a triple mentioned above such as (shape, circle, 0.8).

$$\begin{matrix}
 \begin{bmatrix} shape \\ colour \\ L \\ size \end{bmatrix} & \begin{bmatrix} circle red L N/A \\ square green L small \\ L L L L \\ N/A green L big \end{bmatrix} & \begin{bmatrix} 0.8 & 1 & L & 0 \\ 1 & 0.6 & L & 0.1 \\ L & L & L & L \\ 0 & 0.9 & L & 0.8 \end{bmatrix} \\
 (VM) & (AV) & (WM)
 \end{matrix}$$

It is clear that the concept similarity, which is based on

relevant matrixes and closeness measurement, determines the relations between the concepts. In terms of the closeness measurement, “distances” between the attributes will be calculated. Obviously, the smaller “distance” means much more similar than that of greater “distance”.

In order to calculate the distance, similarity matrixes (also symmetric matrixes, notated as $SM_1, SM_2, SM_j, \dots, SM_m$ for simplicity), where SM_i corresponds to the j^{th} attribute in the AV vector and their different values, are provided. These matrixes show how close these values are (similarity closeness $\in [0, 1]$, in which 1 means perfectly matching, 0 means never matching, other value $\in (0, 1)$ means approximate matching).

Let us recall all of them before we start to discuss our approaches. In order for agents to automatically refine their individual ontologies, firstly, a value matrix (VM), a corresponding attribute vector (AV), and its corresponding weight matrix (WM) must provide. Moreover we suppose that similarity matrixes ($SM_1, SM_2, SM_j, \dots, SM_m$) are available for the proposed ontology and their attributes. In addition, we always treat ontologies are consistent initially. Without this our consistency checking will make no sense.

3. REFINEMENT MECHANISMS

As indicated earlier, ontology evolves overtime in the real world. How to refine ontology on the fly needs to be handled properly. In this section, we consider two primitive operations, i.e. add and delete as the refinement mechanisms based on the ontology structure and cluster analysis discussed earlier in this paper.

3.1 Add Operation

When we consider the add operation, generally there are two cases: (1) a new concept is subsumed by another concept; (2) a new cluster (concept) needs to be created in order to subsume a new concept. It is also possible that a new cluster may generate another bigger new cluster afterwards. Let us consider new concept X (only considering an abstract concept here) in Figure 2. In one case (on the left), X is subsumed by concept B , while in another case (on the right) shows that a new cluster Y (being subsumed by concept A), is generated which subsumes both B and X .

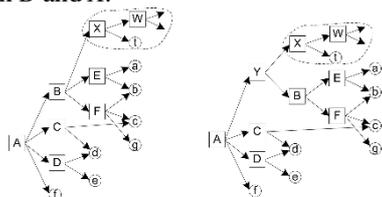


Fig. 2. Two cases of adding operation

For the reason that the number of the attributes might vary from concept to concept, for simplicity, the biggest number will determine the cardinality of the VA . In this

case, there must be some empty elements in VM , here “ N/A ” is used to notate no such a property in VM . Figure 3 is the flow chart to add a new concept through top-down in the hierarchy.

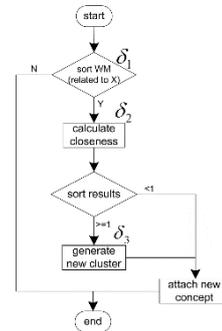


Fig. 3. Flow chart of adding a new concept

First of all, we assume that new concept X has the same attribute vector (AV) as others do. Of course, some properties are “ N/A ” under certain circumstance, so does the weight matrix (WM) with corresponding 0’s value. Additionally, X ’s weight vector X_w and value vector X_v are also provided. The details are as follows.

Step 1: Function *sort WM*

It mainly deals with sorting weight matrix (WM) by referring to X ’s 1’s distribution by referring to its weight vector X_w (each $w_{Xj} \in [0, 1]$). The summary $\sum_{j=1}^m w_{ij} \cdot w_{Xj}$, where $i \in [1, n], j \in [1, m], w_{ij} \in [0, 1]$, corresponding to the element in $WM, w_{Xj} \in [0, 1]$, corresponding to the element in X_w is filtered by a threshold (d_1) which is given a priori.

- (1) The values that are less than the threshold are ignored.
- (2) If all values are less than the given value, the process leads to the *END* of the adding process. That is to say, the new concept X is irrelevant to the existing concept taxonomy structure.
- (3) Otherwise all selected vectors form a new weight matrix *New_WM* and a new value matrix *New_VM* respectively (only related concepts are left after this step).

Step 2: Function *calculate_closeness*

As mentioned before (Section 2.2), SM_j ($j \in [1, m]$) as a similarity matrix (it is symmetric) provides similarity measurements of different values upon a specific attribute. For example, an attribute “*shape*” has a matrix (SM_j) to estimate to what extent the shape “circle” and “ellipse” can be regarded as similar to each other. Figure 4 gives a demonstration. For new concept $X < w_{X1}, w_{X2}, \dots, w_{Xm} >$, calculate $\sum_{j=1}^m w_{Xj} \cdot s_j$

where s_j' coming from above SM_j ($j \in [1, m]$) (Figure 4) based on the New_WM (filtered).

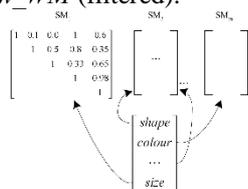


Fig. 4. Similarity matrixes

Step 3: **Function** *sort_result*

Sorting above values in the descending order and selecting candidates (notated as *Selected_New_WM*) whose summary values are greater than the threshold (d_2). If there is no one with a sum value greater than d_2 , go to step 5. Again, getting corresponding *Selecte_New_VM* as well.

Step 4: **Function** *generate_new_cluster*

(1) Appending new concept X 's weight vector X_w and value vector X_v to the *Selected_New_WM* and *Selected_New_VM* respectively.

(2) Selecting all attributes (mapping to AV and satisfying corresponding elements in *Selected_New_WM* which are greater than threshold d_3) to form a New_AV .

(3) Making an intersection to obtain how many attributes are involved (with a corresponding weight in *Selected_New_WM* which is greater than d_3) if possible.

Otherwise go to step 5.

(4) Making an intersection to obtain what values are involved (with a corresponding weight in *Selected_New_WM* which is greater than d_3) if possible.

Otherwise go to step 5.

(5) The newly generated cluster features with value vector Y_v and weight vector Y_w as well. It is worthy noting that the newly generated cluster features with 1's (if it has corresponding attribute) or 0's (no such property) rather than other values between (0, 1).

Step 5: **Function** *attach_new_concept*

Attaching the newly generated concept to the original structure, such as in Figure 2, new concept X (enclosed in a dotted line circle) is attached to original node B if there is no new cluster generated, on the contrary, Y should be attached to node A if new cluster Y is generated while the new concept X is added to the structure.

3.2 Delete Operation

As for deleting a concept from the structure, generally, it should be relatively easy if a simple case (without any sub concepts) is concerned (Figure 5). For other cases, sub concepts of the deleted concept (subgraph) will be attached to its parents (Figure 6).

(1) In Figure 5, suppose the deleted concept is t (on the left), just delete it (see the graph on the right as the result).



Fig. 5. Delete operation— case 1

(2) In Figure 6, suppose the deleted concept is Y (on the left) this time, all its (Y 's) sub concepts (both X and B) will be attached to its parent A (see the graph on the right as the result).



Fig. 6. Delete operation— case 2

Functions such as *generated_new_cluster* and *attach_new_concept* are very useful to explain the add and delete operations. We believe more complex structure can also get some help from approaches here.

4. CONSISTENCY CHECKING

Different partners can contribute to the extension of an ontology. In addition, the same partner may revise the ontology from time to time. Therefore, ontology consistency checking is vital to any applications based on it. Description logics (DLs) as carefully selected parts of first order predicate logics are designed to be expressive enough to be useful for constructing and querying ontologies. Their inference engines are good at answering subsumption and satisfiability queries. In the paper, we use them to check consistency of refined ontology. The basic ideas look as follows.

Generally, C and D denote any concepts, R for the role name. The semantics is given by means of interpretation I . An interpretation is a pair of domain Δ^I and interpretation function \cdot^I , namely $I = (\Delta^I, \cdot^I)$. With parts of constructors, interpretation I for concepts is defined in Table 1

Table 1. Parts of constructors of DLs

Constructor	Syntax	Semantics
Atomic role	R	$R^I \subseteq \Delta^I \times \Delta^I$
Conjunction (AND)	$C \text{ ? } D$	$C^I \mid D^I$
Disjunction (OR)	$C \text{ ? } D$	$C^I \cup D^I$
Negation (NOT)	$\neg C$	$\Delta^I \setminus C^I$

Existential restrict	$?R.C$	$\{x \mid \exists y. \langle x, y \rangle \in R' \wedge y \in C'\}$
Universal restrict	$\forall R.C$	$\{x \mid \forall y. \langle x, y \rangle \in R' \Rightarrow y \in C'\}$

In terms of ontology consistency checking, we have axioms like $\mathcal{Q} D$ (C is subsumed by D , “?” notates subsumption relation in DLs) iff $C ? \neg D ? \perp$, that is to say $\mathcal{Q} \neg D$ is not satisfiable (inconsistent). We use annotations like *property1*, *property2*, etc. to notate different attributes shown in Figure 1 by ignoring what it really means. Based on the DL model, the taxonomy structure (Figure 1) can be expressed more formally as:

- $A ? \exists ? C ? D ? ? \text{property1.f}$
- $B ? A$
- $B ? \exists ? F$
- $C ? A ? ? \text{property2.c} ? ? \text{property3.d}$
- $D ? A ? ? \text{property3.d} ? ? \text{property4.e}$
- $E ? B ? ? \text{property5.a} ? ? \text{property6.b}$
- $F ? B ? ? \text{property6.b} ? ? \text{property2.c} ? ? \text{property7.g}$

For new concept X from Figure 5, it looks like:

- $X ? B$
- $X ? W ? ? \text{property7.t}$

The satisfiability and subsumption can be used during ontology construction to verify whether the ontology is consistent or not. We assume it is consistent initially. When the primitive operations have been processed, DL can answer the questions such as ‘is a concept consistent with the source ontology?’, and ‘Is this concept a part of another?’. For instance (Now look at the case that X is attached to B), $X ? \neg ? \text{property1.f}$ is inconsistent because of $\mathcal{X} B$ (given condition), $\mathcal{B} A$ (already known condition) and which implies $X ? A$. $X ? \neg ? \text{property1.f}$ is not satisfiable because it contradicts with the reasoning result $X ? A$ (*property1.f* is one of properties of concept A shown in Figure 1). As the above consistency checking is conducted according to Figure 1 which is platform- and description language-independent, it is promising to deal with ontology refinement in the is-a hierarchy in general.

5. EXAMPLE

A simple example below will show how a new node is added and a new node is generated by applying cluster analysis. In this example, we assume d_1, d_2 , and d_3 (Figure 3) equal to zero. Euclidean distance, computed as: $\text{distance}(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$, is used to measure “distance”, and the shortest distance in a Euclidean distance matrix will determine which pair of nodes will be fused to form a cluster. First we demonstrate a dendrogram construction through the example, then we con-

sider adding a new node to this hierarchical structure to refine it.

Now suppose there are 4 nodes $V_1(2,3), V_2(5,1), V_3(4,4), V_4(1,2)$. Their corresponding Euclidean distance matrix is shown in Figure 7.a. It shows that the most similar nodes are V_1 and V_4 (the value is 1.4). Then we get a new cluster $V_{14}(1.5,2.5)$, a means of these two nodes is calculated for each dimension. That is to say, after V_1 and V_4 clustered, a new node $V_{14}(1.5,2.5)$ is generated. So now there are three nodes $V_{14}(1.5,2.5), V_2(5,1), V_3(4,4)$. Repeat the above calculation, a revised distance matrix is shown in Figure 7.b. At this stage, the most similar nodes are V_{14} and V_3 (the value is 2.9). Then we get a new cluster, $V_{143}(3.5,4.5)$. Now there are two nodes, $V_{143}(3.5,4.5), V_2(5,1)$. Again we get a revised distance matrix which is shown in Figure 7.c.

The whole process is summarised by a dendrogram as shown in Figure 8 (on the left). Suppose a new node notated as $V_x(6,5)$ is going to be added. As we mentioned before, the entire process works in a top-down approach. The same is here. However, in this example, we only focus on nodes on the first level (the root level namely the top level normally is virtual item annotated as “ENTIRY” or “THING”. In real cases, the attribute matrix will determine to which level the algorithm is affected. The worst case is that all existing concepts will be taken into consideration), namely $V_{143}(3.5,4.5), V_2(5,1)$. At this stage, calculations are between these three nodes. That is $V_x(6,5), V_{143}(3.5,4.5)$, and $V_2(5,1)$. At the end, the distance matrix is shown in Figure 7.d. It is clear that the most similar pair is V_{143} and V_x (the value is 2.5). The revised dendrogram is shown in Figure 8 (on the right). New node V_x is added in and fused with V_{143} to form a cluster V_{x143} which can be regarded as a new generated concept in the ontology refinement process.

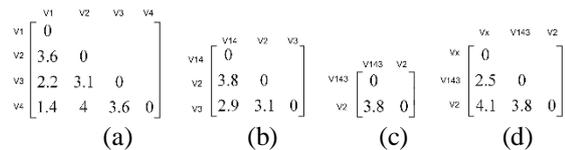


Fig 7. Euclidean distance matrixes

This example demonstrates our ontology refinement process, especially the add operation works well by taking attribute-value pair “distance” and clustering analysis. It is no doubt that in the real world, the refinement for ontology, which includes many concepts, is definitely more complicated than the example. However, the basic ideas and mechanisms to refine the ontology are quite similar to the method here.

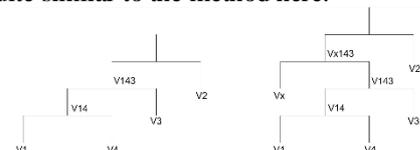


Fig. 8 New cluster in dendrogram construction

6. CONCLUSIONS AND FUTURE WORK

In this paper, our general focus is on developing a new approach to deal with the refinement for ontology evolution in virtual enterprises. A novel clustering algorithm has been applied to classify new concepts. Unlike other methods in ontology refinement, our approaches can classify the new concepts automatically instead of specified or semi-specified manually. In addition, the primitive add and delete operations are provided. Furthermore, we have illustrated that consistency checking should be a crucial element within the ontology evolution life cycle to guarantee that any application has a sound knowledge of the world. The consistency checking of the concepts makes our approach in accordance with knowledge management requirements in preserving knowledge and sharing knowledge. In addition, we believe that clustering analysis is a promising way to knowledge acquisition and ontology management. Moreover, differentiating abstract concepts from concrete ones are useful when we address the hierarchical taxonomic structures.

There are many issues remaining as future work. The semantic meaning of new clustering cannot be fully achieved only through the vector descriptions. Moreover, deleting is not as easy as people first thought. Another aspect is the structure for ontology consistency checking. Consistency checking needs to be emphasised in ontology evolution by considering multiple users. In addition, it is nevertheless essential to address the attribute weighting issue not only in a deterministic environment but also in a non-deterministic environment.

ACKNOWLEDGEMENT

Work reported in this paper is partly supported by Swinburne Vice Chancellor's Strategic Research Initiative Grant 2002-2004 for project "Internet-based e-business ventures".

REFERENCES

[1] Alfonseca, E. and Manandhar, S., Proposal for Evaluating Ontology Refinement Methods, *In: Proceedings of 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, Las Palmas, Spain, 2002. <http://www.ii.uam.es/~ealfon/esp/pubs.html>.

[2] Baader, F., Horrocks, I., and Sattler, U., Description Logics as Ontology Languages for the Semantic Web, *Lecture Notes in Artificial Intelligence. Springer-*

Verlag, 2003. <http://www.cs.man.ac.uk/~horrocks/Publications/>.

[3] Byrne, J. A., Brandt, R., and Bort, O., The Virtual Corporation, *Business Week*, vol. 8, pp. 36-40, February 1993.

[4] Devedzic, V., Understanding Ontological Engineering, *Communications of the ACM*, 45(4), pp. 136-144, April 2002.

[5] Fischer, K., Müller, P. J., Heimig, I., and Scheer, W. A., Intelligent Agents in Virtual Enterprises, *In: Proceedings of the 1st International Conference on Practical Applications of Intelligent Agents and Multi-Agent Technology (PAAM'96)*, London, 1996.

[6] Gruber, T. R., Toward Principles for the Design of Ontologies Used for Knowledge Sharing, KSL-93-04, Knowledge Systems Laboratory, Stanford University. <http://ksl-web.stanford.edu/>.

[7] Horrocks, I. and Sattler, U., A Description Logic with Transitive and Inverse Roles and Role Hierarchies, *Journal of Logic and Computation*, 9(3), pp. 385-410, 1999.

[8] Maedche, A. *Ontology Learning for the Semantic Web*, Kluwer Academic Publishers, 2002.

[9] McNeill, F., Bundy, A., and Schorlemmer, M., Dynamic Ontology Refinement, Informatics Research Report EDI-INF-RR-0177, School of Informatics, University of Edinburgh, June 2003.

[10] Noy, N. F. and Klein, M., Ontology Evolution: Not the Same as Schema Evolution, *Knowledge and Information Systems*, In press. Available as SMI technical report SMI-2002-0926 (2002). <http://smi.stanford.edu/people/noy/publications.html>.

[11] Noy, N. F. and Musen, M. A., PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment, *In: Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-2000)*, Austin, TX, 2000.

[12] Petersen, S. A., Using Agents to Support the Selection of Virtual Enterprise Teams, *In: Proceedings of 4th International Bi-Conference Workshop on Agent-Oriented Information Systems (AOIS-2002)* (at AAMAS 2002), Bologna, Italy, July 2002.

[13] Petersen, S. A., Divitini, M., and Matskin, M., An Agent-based Approach to Modelling Virtual Enterprises, *International Journal of Production Planning and Control*, 12(3), pp. 224-233, April 2001.

[14] Stojanovic, L., Maedche, A., Motik, B., and Stojanovic, N., User-driven Ontology Evolution Management, *In: Proceedings of the 13th European Conference on Knowledge Engineering and Knowledge Management EKAW*, Madrid, Spain, 2002.

[15] Tryon, R. C. and Bailey, D. E. *Cluster Analysis*, McGraw-Hill, New York, NY, 1973.

Risk Evaluation for Virtual Enterprise

Min Huang¹, Hongmei Yang¹, Xingwei Wang¹, W.H.Ip², Dingwei Wang¹

¹ Box 135, Faculty of Information Science and Engineering, Northeastern University,
Shenyang, Liaoning 110004, China, mhuang@mail.neu.edu.cn

² Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University,
Hung Hom, Hong Kong, China

ABSTRACT

Virtual Enterprise is the potential mode of enterprise in the future. The risk management for virtual enterprise is the new research area recently. In virtual enterprise, the enterprise operation is always organized by project mode and there is always less historical data and there are many uncertain factors. Hence, in this paper, the fuzzy synthetic evaluation model for the risk evaluation of virtual enterprise is established focus on the project mode and uncertain characteristics of virtual enterprise. In the 5 levels model, the goal and sub-goal of the enterprise, the process of the project, as well as the risk event and risk factors are considered. The case study suggests that the method is useful.

Keywords: virtual enterprise, risk evaluation, fuzzy synthetic evaluation

1. INTRODUCTION

Risk Evaluation, a significant stage in risk management, is defined as a process in which the acceptance level of integrated enterprise risk is determined based on single-risk evaluation. Common evaluating methods include balance-point method, sensitive analysis and probability analysis, etc. [1-4], in which the simple relation between one kind of profit and its relevant parameter is discovered from a standpoint of profit then described mathematically or probabilistically. However, since an enterprise, especially a virtual enterprise is a complicated system consisting of a number of functional elements, its risks should be analyzed and evaluated from a systematic point of view so that the general objective of the system and functions of all parts can be managed and the relations in-between acknowledged. Only when the operating regularity of each kind of risk is studied can the integrated risk level be determined. Besides, since a virtual enterprise is a dynamic alliance [5-7] which exists temporally with the aims of grasping some new opportunity in the market, two main features of it is considered in this research. First, the operation of VE is always organized by project mode, so the integrated risk level is obtained based on the evaluation of each sub-procedure risk level in the analysis. Second, there is neither basic data nor objective probabilistic distribution for reference but personal experiences and subjective judgment, which is of great fuzziness. As a result, fuzzy method is proposed in our research. In this research, a risk evaluation model for virtual enterprises is established based on fuzzy mathematic theory from a systematic point of view and case study provided to prove its effectiveness.

2. FUZZY SYNTHETIC EVALUATION THEORY

Comparison is often needed for judging in productive operation, scientific researches and daily life. To

evaluate is easy if there is only one factor to be considered. However, it is always necessary to judge synthetically because there are always a number of attributive factors reflecting different characteristics, especially when dealing with complicated systems. If fuzzy features are also involved, it becomes a fuzzy synthetic evaluation problem[8-11].

2.1 Single-level Fuzzy Synthetic Evaluation

The single-level synthetic evaluation is used when the factor to be considered is not too much. Four factors need to be considered in this process:

- 1) Factor set of all factors to consider in the synthetic evaluation, denoted by the set

$$U = \{u_1, u_2, \dots, u_m\} \quad (1)$$

- 2) Fuzzy subsets indicating importance levels of factors in the synthetic evaluation, denoted by

$$A = \frac{a_1}{u_1} + \frac{a_2}{u_2} + \dots + \frac{a_m}{u_m} \quad (2)$$

among which a_1, a_2, \dots, a_m represents importance levels of the corresponding factors.

- 3) Comment set of evaluation result, denoted by

$$V = \{v_1, v_2, \dots, v_m\} \quad (3)$$

- 4) Evaluating matrix of fuzzy relation between factor domain and comment domain, denoted by :

$$R = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_m \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{pmatrix} \quad (4)$$

where r_{ij} represents membership degree of object to be v_j from the standpoint of u_i . R_i represents the

mono-factor evaluation to u_i . It is a fuzzy subset of V .

The fuzzy matrix being compound-calculated, then:

$$A \circ R = B \tag{5}$$

or

$$(a_1, a_2, \dots, a_m) \circ \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \text{L} & \text{L} & & \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{pmatrix} = (b_1, b_2, \dots, b_n) \tag{6}$$

Let

$$B = (b_1, b_2, \dots, b_n) \tag{7}$$

B is a fuzzy subset of comment domain V , indicating the membership degree of evaluated results to comment set, which is usually determined according to maximum membership degree principle.

A number of methods can be applied to compound calculation in fuzzy matrix, of which the two most commonly-used types are:

- 1) Main-factor-determining type

$$b_j = \max[\min(a_1, r_{1j}), \min(a_2, r_{2j}), \dots, \min(a_n, r_{nj})] \tag{8}$$

It is best applied to the situation of which the optimum is determined by one optimal single factor.

- 2) Weighted-mean-determining type

$$b_j = \sum_{i=1}^m a_i \cdot r_{ij} \tag{9}$$

It takes function of every factor in general evaluation into account, giving attention to the influence of each, which makes it the most suitable for situations that require integrated evaluative index.

The second type is used in our analysis.

2.2 Multi-level Fuzzy Synthetic Evaluation (MFSE)

A lot of factors should be considered in complicated problems and systems. These factors might belong to different categories and hierarchies, which are in need of a multi-level evaluation: firstly they are primarily and synthetically evaluated within lowest hierarchy; then a higher hierarchy synthetic evaluation is given based on the results of its prior step in lower hierarchy; and so on, till the evaluation of top hierarchy is gotten. This method can specify the status and functions of all factors in the general evaluation and adopt all information derived from the factors.

Let L be the level number of the MFSE model, l be the counter of level. The integrated level is on the level 0, the sub-goal is on the level 1, and so on. Let K_l be the factor number of level l , k be the counter of factor number, $a_k = (a_{k1}, a_{k2}, \dots, a_{kM})$ be the weight

vector of factors under factor k on level l , M be the factor number under factor k on level l , m be factor counter under factor k on level l , $V = (v_1, v_2, \dots, v_n)$ be the set of evaluation, b_k be the fuzzy evaluation of factor k on level l , R_k be the fuzzy evaluation matrix composed of fuzzy evaluation of each factor under factor k on level l .

The steps of MFSE are as follows.

Step 1: $l=L-2$; k from 1 to K_l , determine R_k according to V , then:

$$b_k = a_k \circ R_k = (b_{k1}, b_{k2}, \dots, b_{kn}) \quad (k=1, 2, \dots, K_l) \tag{10}$$

The synthetic evaluation of level l is finished and turns to step 2.

Step 2: If $l=0$ stop. Otherwise, let $l=l-1$ and turn to step 3.

Step 3: Let k from 1 to K_l , according to $b_m = (b_{m1}, b_{m2}, \dots, b_{mn}) \quad (m=1, 2, \dots, M)$, R_k is given as follows:

$$R_k = \begin{pmatrix} B_1 \\ B_2 \\ \text{M} \\ B_M \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ & \text{L} & \text{L} & \\ b_{M1} & b_{M2} & \dots & b_{Mn} \end{pmatrix} \tag{11}$$

Then, according to equation (10) the synthetic evaluation of level l is finished, and turns to step 2.

3. FUZZY SYNTHETIC EVALUATION MODEL OF RISK FOR VIRTUAL ENTERPRISE

Risk evaluation is performed based on risk identification and risk estimation. Such information can be acquired in risk identification as:

- a) General Objective and sub-objectives, the contents of which do not overlap;
- b) Components or operation procedures of the enterprise, the contents of which do not overlap;
- c) Whether the procedures and the objectives are relevant;
- d) Every possible risk;

While risk estimation provides:

- a) Significance of each sub-objective and its weight;
- b) Importance of relevant procedures or components to the sub-objectives and their weights;
- c) Risk allocation, e.g., which procedure or component the risks affect, to which objective they are

- related;
- d) Significance of each risk and its weight;
- e) Factor-factor relations within risks and the weights of factors;
- f) Fuzzy description of risk factors.

Two factors are determining the significance of a risk: risk probability and risk lost. Criterion of assessment^[3] is listed as follows for analyzing the two factors (see Table 1 and 2, from which a fuzzy description of risk factors can be derived.

Based on risk identification and estimation, considering the project organization mode and the uncertain characteristics of VE, the fuzzy synthetic evaluation

hierarchical model (FSEHM) of risk evaluation is given in Figure 1. In this FSEHM model, each process of the VE project is considered, as well as the fuzzy description of each risk factor.

In the FSEHM model, there are 5 levels from level 0 (top) to level 4 (bottom). The level 0 is the general object of VE to minimize risk. Table 3 provides the criterion for evaluating general objective. The level 1 is the sub-objectives the VE pursued, which are different for a special VE. The level 2 is the procedures, which are the activities in the VE project. The level 3 are the risk events, which will cause the risk in the procedures for different sub-objective. The level 4 are the risk factor of the risk event. As mentioned above, two risk factors, risk probability and risk loss, are considered for each risk event.

Table 1 Rank criterion for risk probability

Severity	Incidental possibility description	Criterion for evaluation
0	Nil	0%—5%
1	Slight	5%—15%
2	Lower	15%—25%
3	Lower	25%—40%
4	Mediat	40%—60%
5	A bit high	60%—75%
6	High	75%—85%
7	Higher	85%—95%
8	Highest	95%—100%

Table 2 Rank criterion for risk lost

Severity	Lost description	Criterion for Evaluation
0	Nil	No influence on producing system
1	Slight	Slight influence but almost no product demands reproduction.
2	Lower	Producing system influenced,part of products demand reproduction
3	Lower	Producing system is operatable with a deteriorating capacity. Half products demand reproduction.
4	Mediat	Operatable producing system with a sharply-decreasing capacity, majoroty of products demand reproduction
5	A bit high	Producing system operates abnormally,no significant segment paralysed, reparable
6	High	Producing system operates abnormally,significant segment paralysed,reparable
7	Higher	Producing system damaged severely, almost disrepairable
8	Highest	Producing system breaks down , disrepairable and may endanger operator's life

4. CASE STUDY

The example involves the real life problem of an enterprise, which bid for a market opportunity for the manufacturing lamp. The project consists of 5 processes, the precedence relationship represented by the Activity-on-Arc mode is shown in Figure 2. The owner has the ability of design and core manufacturing, while the bulb manufacturing, cap manufacturing, and assembly processes are finished by partner.

The objective of the VE is to minimize the general risk; the sub-objectives are to minimize cost risk, coordination risk, the time risk and the quality risk respectively. The weight of each sub-objective to the subjective is shown in table 4. The weight (W) of each procedure (P) to sub-objective (S) is shown table 5. The risk events are shown in table 6. Their relationship to each procedure under a sub-objective and the weights to the corresponding procedure are show in table 7. The weight of probability and loss to each risk event is shown in Table 6. The fuzzy description of probability and loss for each risk event is show in Table 8.

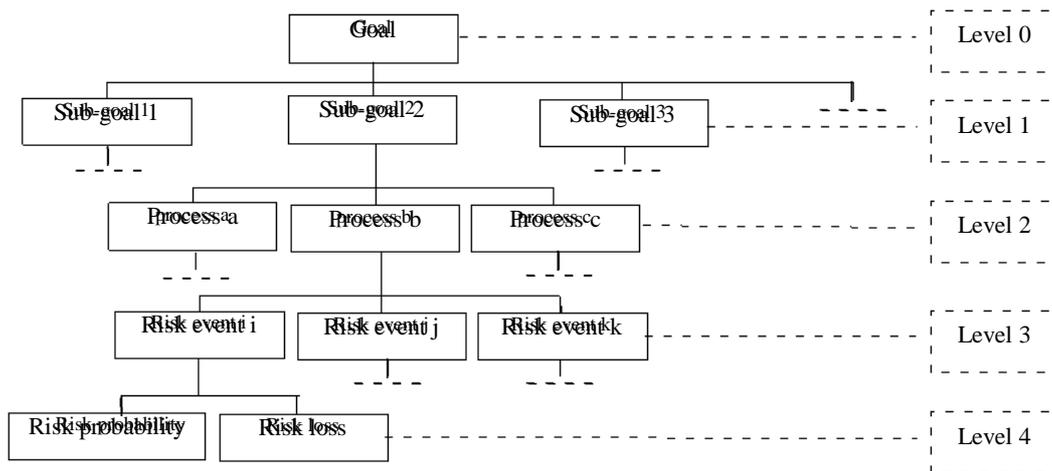


Figure 11 The hierarchical model for risk evaluation

Table 3 Integrated evaluation standard for risk level

Risk level	0	1	2	3	4	5	6	7	8
Description	Nil	Slight	Lower	Low	Mediate	A bit high	High	Higher	Highest

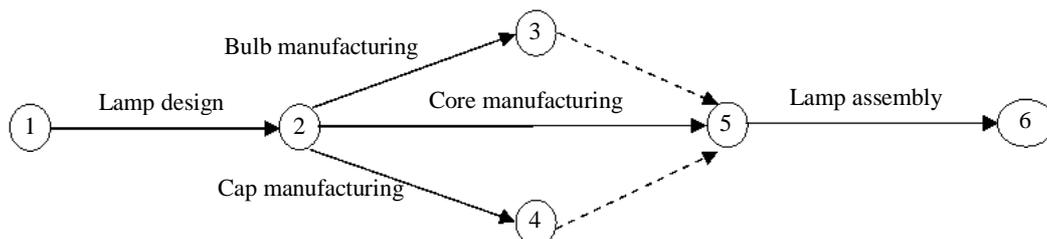


Figure 2 The network for the lamp manufacturing

Table 4 The weight of the sub-goals to the goal

Sub-goal	Cost	Coordination	Time	Quality
Weight	0.3	0.3	0.15	0.25

Table 5 The weight of the processes to the sub-goals

W \ S \ P	Process 1	Process 2	Process 3	Process 4	Process 5
Sub-goal1	0.5	0.3	0.05	0.15	0
Sub-goal2	0.2	0.5	0.1	0.1	0.1
Sub-goal3	0	0.2	0.3	0.3	0.2
Sub-goal4	0	0	0.4	0.6	0

Table 9 shows the result of risk evaluation for the VE. According to the principle of maximum membership and the method of fuzzy synthetic evaluation, the result of integrated risk evaluation is 0.234, indicating a grade 3 risk.

5. CONCLUSION

In accordance with project organization mode and the

great uncertainties and information fuzziness of virtual enterprise, a risk evaluation model is established in this paper based on procedures in project network and the theory of fuzzy synthetic evaluation, and the integrated consideration of influences of different factors. Examples have shown its simple, operability and charity for comprehension. The result reflects the integrated risk level effectively.

Table 6 The risk events and probability and loss weight of them

Risk event No.	Risk event	Risk probability	Risk loss
1	Design method	0.4	0.6
2	Designer's level	0.5	0.5
3	Delay of bulb	0.4	0.6
4	Bad rate of core	0.35	0.65
5	Delay of cap	0.5	0.5
6	Communication with partner	0.35	0.65
7	Selection of bulb partner	0.8	0.2
8	Contract award	0.15	0.85
9	Strategy of partner	0.25	0.75
10	Selection of cap partner	0.3	0.7
11	Selection of assembly partner	0.7	0.3
12	The experience of the designer	0.5	0.5
13	The complexity of the product	0.6	0.4
14	Contract management	0.8	0.2
15	The capability of the enterprise	0.3	0.7
16	The experience of the worker	0.7	0.3
17	The capability of assembly partner	0.5	0.5
18	The reputation of bulb partner	0.2	0.8
19	The reputation of cap partner	0.65	0.35
20	The reputation of assembly partner	0.5	0.5

Table 7 The relationship and weight of risk event to the processes under the sub-goals

D \ P S	Process 1		Process 2	Process 3	Process 4	Process 5
	Sub-goal1	Risk 1 (0.7) Risk 2 (0.3)	Risk 3 (1.0)	Risk 4 (1.0)	Risk 5 (1.0)	-
Sub-goal2	Risk 6 (1.0)	Risk 7 (0.5) Risk 8 (0.4) Risk 9 (0.1)	-	Risk 10 (1.0)	Risk 11 (1.0)	
Sub-goal3	Risk 12 (0.2) Risk 13 (0.8)	Risk 14 (1.0)	Risk 15 (0.8) Risk 16 (0.2)	-	Risk 17 (1.0)	
Sub-goal4	-	Risk 18 (1.0)	-	Risk 19 (1.0)	Risk 20 (1.0)	

* P: process; S: sub-goal; D: relationship and weight of risk event to the processes under the sub-goals.

ACKNOWLEDGEMENTS

The authors wish to thank the support of the National Natural Science Foundation of China (Project no. 70101006, 60473089, 60003006), Liaoning Province Natural Science Foundation of China (Project no. 20032019, 001015), the Scientific Research Foundation for the Returned Overseas Chinese Scholars by SEM of China, Modern Distance Education Engineering Project by MoE of China, and the Research Grants Council of Hong Kong, China (Project no. PolyU 5167/99E).

REFERENCES

- [1] Kliem, R.L. and Ludin, I.S., *Reducing Project Risk*, England: Gower Publishing Limited Gower House, 1997.
- [2] Bier, V.M., Haimes, Y.Y., Lambert, J.H., Matalas, N.C. and Zimmerman, R., "Survey of approaches for assessing and managing the risk of extremes", *Risk Analysis*, Vol. 19, No. 1, pp 83-94, 1999.
- [3] Deng Z.Y., *Strategies for Risk Profit Management*, Hong Kong: Wan Li Organization, Wan Li Bookshop, 1993.
- [4] Lu Y.J. and Lu J.Y., *Project Risk Management*, Beijing: Tsinghua University Press, 1998.
- [5] Park K. H. and Favrel J., "Virtual enterprise - information system and networking solution", *Computers and Industrial Engineering*, Vol. 37, No. 1, pp441-444, 1999.
- [6] Chalmeta, R. and Grangel, R., "ARDIN extension for virtual enterprise integration", *The Journal of Systems and Software*, Vol. 67, No. 1, pp141-152, 2003.
- [7] Jiang Y., Fan Y. and Wu C., "Research into Dynamic Association and Agility of Enterprise", *Theory and Practice of Systems Engineering*, Vol. 22, No.1, pp71-74, 2002.
- [8] Hwang, C.L. and Yoon, K., *Multiple Attribute Decision Making: Methods and Applications*, Berlin: Springer-Verlag, 1981.
- [9] Buckley J.J., "Fuzzy hierarchical analysis", *Fuzzy Sets and Systems*, Vol. 17, No. 1, pp233-247, 1985.
- [10] Wang F., *Fuzzy Mathematics and Science of Engineering*, Harbin: Harbin Marine Industry College

Press, 1998.

[11] Chen S.Y., *Fuzzy-sets Theory and Application for*

Engineering, Beijing: Beijing National Defense Industry Press, 1998.

Table 8 The fuzzy description of probability and loss for each risk event with no risk control

E	D	R	0	1	2	3	4	5	6	7	8
		Probability	0.0	0.0	0.0	0.3	0.5	0.2	0.0	0.0	0.0
1	Loss	0.0	0.0	0.0	0.0	0.2	0.8	0.0	0.0	0.0	
	Probability	0.0	0.0	0.3	0.3	0.3	0.1	0.0	0.0	0.0	
2	Loss	0.0	0.0	0.0	0.0	0.3	0.5	0.2	0.0	0.0	
	Probability	0.0	0.0	0.1	0.2	0.4	0.3	0.0	0.0	0.0	
3	Loss	0.0	0.3	0.4	0.3	0.0	0.0	0.0	0.0	0.0	
	Probability	0.0	0.0	0.0	0.0	0.0	0.6	0.4	0.0	0.0	
4	Loss	0.0	0.0	0.0	0.2	0.6	0.2	0.0	0.0	0.0	
	Probability	0.0	0.0	0.1	0.2	0.4	0.3	0.0	0.0	0.0	
5	Loss	0.0	0.3	0.4	0.3	0.0	0.0	0.0	0.0	0.0	
	Probability	0.0	0.0	0.0	0.5	0.5	0.0	0.0	0.0	0.0	
6	Loss	0.0	0.0	0.0	0.5	0.5	0.0	0.0	0.0	0.0	
	Probability	0.0	0.0	0.0	0.3	0.4	0.3	0.0	0.0	0.0	
7	Loss	0.0	0.0	0.1	0.2	0.3	0.3	0.1	0.0	0.0	
	Probability	0.0	0.0	0.1	0.8	0.1	0.0	0.0	0.0	0.0	
8	Loss	0.0	0.0	0.1	0.8	0.1	0.0	0.0	0.0	0.0	
	Probability	0.0	0.5	0.5	0.0	0.0	0.0	0.0	0.0	0.0	
9	Loss	0.0	0.5	0.	0.0	0.0	0.0	0.0	0.0	0.0	
	Probability	0.0	0.0	0.0	0.0	0.3	0.4	0.3	0.0	0.0	
10	Loss	0.0	0.8	0.2	0.0	0.0	0.0	0.0	0.0	0.0	
	Probability	0.0	0.1	0.3	0.5	0.1	0.0	0.0	0.0	0.0	
11	Loss	0.0	0.0	0.0	0.4	0.4	0.2	0.0	0.0	0.0	
	Probability	0.0	0.0	0.5	0.5	0.0	0.0	0.0	0.0	0.0	
12	Loss	0.0	0.0	0.0	0.3	0.5	0.2	0.0	0.0	0.0	
	Probability	0.0	0.0	0.4	0.3	0.3	0.0	0.0	0.0	0.0	
13	Loss	0.0	0.0	0.6	0.2	0.2	0.0	0.0	0.0	0.0	
	Probability	0.0	0.0	0.0	0.5	0.5	0.0	0.0	0.0	0.0	
14	Loss	0.0	0.0	0.0	0.5	0.5	0.0	0.0	0.0	0.0	
	Probability	0.0	0.0	0.4	0.3	0.3	0.0	0.0	0.0	0.0	
15	Loss	0.0	0.0	0.0	0.4	0.4	0.2	0.0	0.0	0.0	
	Probability	0.0	0.0	0.3	0.3	0.3	0.1	0.0	0.0	0.0	
16	Loss	0.0	0.0	0.0	0.0	0.3	0.5	0.2	0.0	0.0	
	Probability	0.0	0.0	0.0	0.0	0.0	0.2	0.4	0.4	0.0	
17	Loss	0.0	0.0	0.5	0.5	0.0	0.0	0.0	0.0	0.0	
	Probability	0.0	0.0	0.3	0.3	0.2	0.2	0.0	0.0	0.0	
18	Loss	0.0	0.0	0.0	0.2	0.2	0.2	0.4	0.0	0.0	
	Probability	0.0	0.0	0.0	0.0	0.0	0.6	0.4	0.0	0.0	
19	Loss	0.0	0.0	0.00	0.2	0.6	0.2	0.0	0.0	0.0	
	Probability	0.0	0.8	0.2	0.0	0.0	0.0	0.0	0.0	0.0	
20	Loss	0.0	0.0	0.0	0.0	0.3	0.5	0.2	0.0	0.0	

* E: risk event; R: risk level; D: fuzzy description.

Table 9 Result of risk evaluation for the VE

Risk level	0	1	2	3	4	5	6	7	8
Membership degree	0.074	0.081	0.100	0.234	0.212	0.209	0.070	0.020	0.00

Risk Sorting for Enterprise under EC Environments

Min Huang¹, Fenge Li¹, Xingwei Wang¹, W.H.Ip²

¹ Box 135, Faculty of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, China, mhuang@mail.neu.edu.cn

² Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China

ABSTRACT

With the rapid development of internet and emerging of global economic, risk management for enterprise under EC (Electronic Commerce) environments has drawn attentions of many researchers. In this paper, the characteristics of risk for EC enterprise are analyzed. Further, focused on the project organization mode and the uncertain factor of the enterprise under EC, which are main different characteristics from the conventional enterprise, enterprise risk sorting, which is one of the key problems of risk management under EC environments, is studied by using fuzzy ISODATA cluster method based on fuzzy describing of risks. Case study suggests the effectiveness of the method.

Keywords: electronic commerce, risk management, risk sorting, fuzzy cluster

1. INTRODUCTION

As the developing of computer and communication technology and the forming of global market economy, internet-oriented EC (electronic commerce) has become a trend and a must for enterprise operation [1,2], which made enterprise management under EC environments a popular issue among the field of management [3-6]. Since there are more uncertainties and unstable factors in enterprise operation under EC environments, such as the dynamic nature of enterprise alliance in different districts, regions even nations, and the leading function of enterprise development caused by new technology progress, enterprises are facing a larger number of conspicuous risks and enterprise risk management, as a special kind of enterprise management, is becoming more significant.

Risk sorting, one of the fundamental works of risk management, aims at acquiring overall and profound knowledge of all kinds of risks after sorting the existed risks from different point of view and according to different criteria and therefore suggests a well-aimed management. Many researches has done about risk sorting [7], however, most of them are qualitative method [8,9], and the features of enterprise under EC are not considered. In this paper, we focus on the project organization mode and the uncertain factor of the enterprise under EC, which are main difference from the conventional enterprise. In the research risks are sorted by the sub-procedures of enterprise projects under EC environments. With the focus on the uncertain features of enterprise risks under EC environments, a fuzzy risk description mechanism is established and a fuzzy sorting is given to the risks by fuzzy ISODATA [10,11] cluster method, which providing a satisfactory preparation for risk evaluation and control.

2. FUZZY RISK DESCRIPTION MECHANISM

The risks to be sorted are denoted by set

$A = \{A_1, A_2, \dots, A_n\}$, factor set to be considered by $B = \{B_1, B_2, \dots, B_m\}$. n represents the number of risks, m represents the number of evaluation factors, hence the descriptive vector of each risk A_i is:

$$X_i = (x_{i1}, x_{i2}, \dots, x_{im}) \quad (1)$$

where, x_{ij} is the evaluation of i risk by j factor.

Then, $X_{m \times n} = \{X_1, X_2, \dots, X_n\}$ is the factor evaluation matrix.

Let $C = \{C_1, C_2, \dots, C_c\}$ be the cluster set, which is sub-procedures of the enterprise project. c is the number of sorting, which is equal to the number of sub-procedures of the enterprise project. Let $V_{c \times m}$ be a group of cluster central vector, and then V_h be cluster center for sort h .

The aim of risk sorting is to get the sorting matrix $R_{c \times n}$ for common sorting and the fuzzy sorting matrix $\tilde{R}_{c \times n}$ for fuzzy sorting, where, $R \in M_c$

$$r_{hj} = \begin{cases} 1 & i \in \text{sort } h \\ 0 & i \notin \text{sort } h \end{cases} \quad (2)$$

And $\tilde{R} \in M_{fc}$, $M_{fc} = \{\tilde{R} \in V_{c \times n} \mid |r_{hi}| \in [0, 1], \forall h, i\}$;

$$\sum_{h=1}^c r_{hi} = 1, \forall i; \sum_{i=1}^n r_{hi} > 0, \forall h \quad (3)$$

3. FUZZY ISODATA CLUSTER ANALYSIS

Fuzzy ISODATA cluster is a fuzzy sorting method based on fuzzy descriptions [10,11].

The cluster criterion for common sorting aims at minimizing the object function:

$$J(R, V) = \sum_{i=1}^n \sum_{h=1}^c r_{hi} \|x_i - V_h\|^2 \quad (4)$$

where $\|x_i - V_h\|$ represents the distance between sample x_i and cluster central vector V_h .

The sorting is to minimize the object function by determining an appropriate matrix $R_{c \times n}$ and $V_{c \times m}$

The cluster criterion for fuzzy sorting aims at minimizing the object function:

$$J(R, V) = \sum_{i=1}^n \sum_{h=1}^c (r_{hi})^q \|x_i - V_h\|^2 \quad (5)$$

where q is a parameter equals or greater than 1. If $q=1$ and $r_{hi} \in \{0, 1\}$, function (5) transforms to a common sorting.

This criterion minimizes the object function by determining an appropriate matrix $R_{c \times n}$ and $V_{c \times m}$ group of cluster central vectors.

3.1 Fuzzy ISODATA Cluster Steps

1) Data normalization: using standardizing method

$$x'_{ik} = \frac{x_{ik} - \bar{x}_i}{\sigma_i} \quad (6)$$

In equation (6):

$$\bar{x}_i = \frac{1}{m} \sum_{k=1}^m x_{ik}, \sigma_i = \sqrt{\frac{1}{m-1} \sum_{k=1}^m (x_{ik} - \bar{x}_i)^2} \quad (7)$$

2) Determine the number of cluster c , $2 \leq c \leq n$, and an initial fuzzy sorting matrix $R^{(0)} \in M_{fc}$, let $l=0$, go to 3).

3) Calculate the cluster central vector for $R^{(l)}$:

$$V^0 = (V_1^0, V_2^0, \dots, V_c^0)^T \quad (8)$$

In equation (8):

$$V_h^0 = \frac{\sum_{i=1}^n (r_{hi})^q x_i}{\sum_{i=1}^n (r_{hi})^q} \quad (9)$$

q is a positive number greater than 1. Go to 4).

4) Modify fuzzy sorting matrix $R^{(0)}$

$$r_{hi}^{(l+1)} = \frac{1}{\sum_{t=1}^c \left(\frac{\|x_i - V_h^{(l)}\|}{\|x_i - V_t^{(l)}\|} \right)^{\frac{1}{q}}}, \quad (\forall h, i) \quad (10)$$

Go to 5).

5) Compare $R^{(l)}$ and $R^{(l+1)}$, if for the given error accuracy $\epsilon > 0$ have:

$$\max_i |r_{hi}^{(l+1)} - r_{hi}^{(l)}| \leq \epsilon \quad (11)$$

$R^{(l+1)}$ and $V^{(l)}$ are the results wanted. Stop iterating. Otherwise, $l = l + 1$, back to 3).

3.2 Evaluation of Fuzzy ISODATA Cluster

Sorting Coefficient Index (SCI) can be applied to the evaluation the results of the fuzzy cluster:

$$F_c(R) = \frac{1}{n} \sum_{i=1}^n \sum_{h=1}^c r_{hi}^2 \quad (12)$$

where $R \in M_{fc}$, $F_c(R) = 1$.

Hence as $F_c(R)$ approaches 1, fuzziness of the ultimate sorting decreases, which indicates a better outcome.

4. CASE STUDY

4.1 Case Description

Let set A, which denotes the risks to be sorted, be:

A = {demand risk, competition risk, overflow risk, market fluctuation risk, economic descent risk, financial risk, environmental risk, policy risk, quality risk, cost risk, time risk, technical risk, communicative risk, technical up-link risk, technical out-leak risk, data quality risk, information system security risk, organization and management risk, credit risk, mobility risk, excitation risk, strategic supply lost risk, investment hold-up risk, investment attainment risk}

It can be seen that $n=24$. Two situations are considered here, the number of sorting and factors for them are respectively:

$c_1 = 11, m_1 = 40$ and $c_2 = 17, m_2 = 45$.

4.2 Simulation Analysis

Two methods have been tried out here:

Method 1: Cluster on the basis of given initiative fuzzy sorting matrix $R^{(0)}$.

Method 2: Cluster on the basis of given fuzzy cluster unstable influence on cluster outcome. Therefore, the central vector $V^{(0)}$.

Number of sorting c , the initiative fuzzy sorting matrix $R^{(0)}$, error accuracy ϵ and parameter q might affect the cluster results. So, the analyses of them are given in the following subsection.

(1) Influences of Different Methods on Cluster

Let error accuracy $\epsilon = 0.00001$, parameter $q = 1.45$. The clusters of the two methods are compared in Table1.

As in Table 1, method 1 shows a superior outcome to method 2 because it takes relative constraints into consideration when choosing the initiative fuzzy sorting matrix.

Table 1 Comparison of the two methods

SCI	Method	Method 1	Method 2
No. of sorting			
$c_1 = 11$		0.601126	0.243985
$c_2 = 17$		0.780527	0.101113

(2) Influences of Different Initiative Fuzzy Sorting Matrix on Cluster

Figure 1 shows the graphs of the SCI against evaluation of initiative sorting matrix (error accuracy $\epsilon = 0.00001$, parameter $q = 1.45$).

In Figure 1, initiative sorting matrix shows a very

actual risks and their characteristics should be given major consideration when determine the initial fuzzy sorting matrix.

(3) Influence of Different Error Accuracy on Cluster

Figure 2 shows the curves of cluster outcome against error accuracy ϵ (parameter $q = 1.45$).

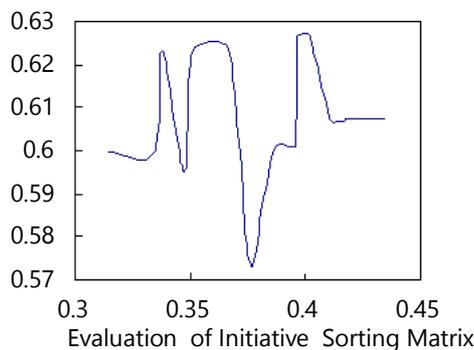
Figure 2 indicates that an error accuracy ϵ of 0.01 or less than 0.01 leads to a satisfactory outcome. This complies with that rule that cluster outcome improves as the error accuracy increases while deteriorates as it decreases.

(4) Influence of Different Parameter q on Cluster

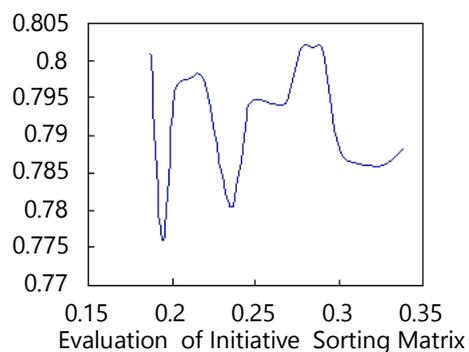
The curves of cluster outcome against parameter q are represented by Figure 3.

It can be concluded from Figure 3 that an increasing parameter q deteriorates the cluster. The fuzziness of sorting becomes smaller as parameter q gets closer to 1. When q equals 1 the sorting becomes a solid sorting. When q increases the sorting becomes fuzzier, which leads to a less definite implication. $q = 1.45$ is chosen according to the actual situation.

The analysis above also indicates that the numbers of sorting does not show an obvious influence on cluster outcome. Nevertheless, more operational time of the procedure is needed as it increases: when $c = 11$, the operational time is 3 seconds; when $c = 17$, it is 4 seconds.



(a) $c=11$



(b) $c=17$

Figure 1 Curves of cluster outcome against initiative sorting matrix

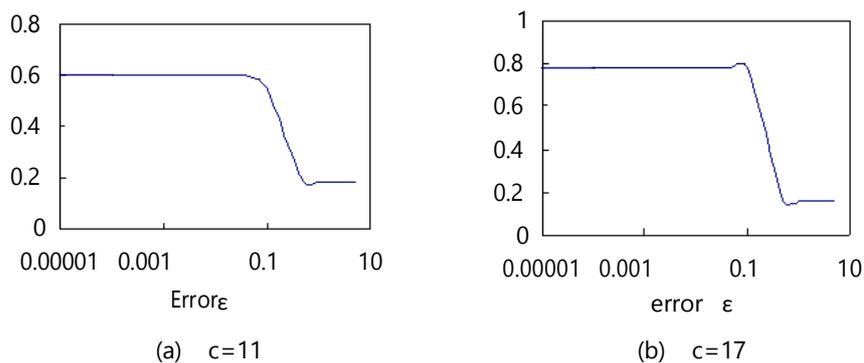


Figure 2 Curves of cluster outcome against error accuracy

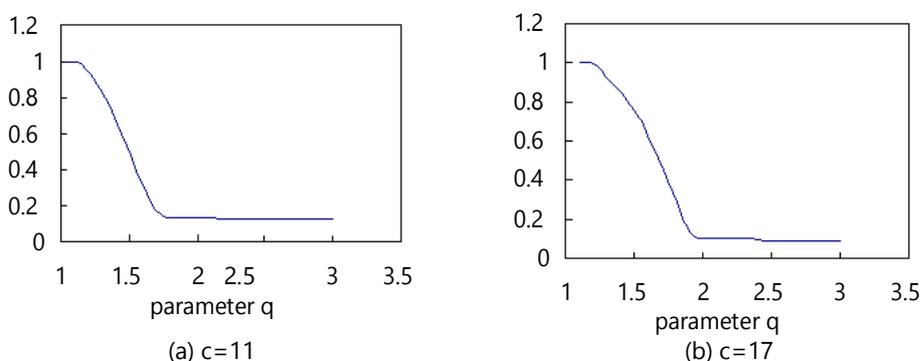


Figure 3 Curves of cluster outcome against parameter

(5) Sorting Outcomes

According to the analysis from (1) to (4), the ultimate sorting outcomes are listed in Table 2 when the number of sorting $c = 11$, error accuracy $\epsilon = 0.00001$, parameter $q = 1.45$, evaluation of estimate of initiative sorting matrix is 0.386458.

It can also be seen in Table 2 that the sorting matrix meets the 3 requirements of fuzzy space division M_{fc} throughout the iteration because the sample initiative sorting matrix has been normalized. It shows:

$$r_{hi} \in [0,1], \forall h,i ; \sum_{h=1}^c r_{hi} = 1, \forall i ; \sum_{i=1}^n r_{hi} > 0, \forall h$$

Hence the model is proved correct and the sorting result complies with objective facts as in Table 2.

5. CONCLUSION

In this research enterprise risk sorting – the key issue of enterprise risk management under EC environment – has been studied focus on the project organization mode and

uncertain factors of the enterprise under EC. A fuzzy description system to enterprise risk is established firstly in accordance with the uncertain features of enterprise risk under EC environments then enterprises risk sorted based on the method of fuzzy ISODATA cluster considering the project sub-procedures is present. The case study involved has analyzed influences of parameters on cluster outcomes, which proves the effectiveness of the method and its significant guidance for enterprise risk management.

ACKNOWLEDGEMENTS

The authors wish to thank the support of the National Natural Science Foundation of China (Project no. 70101006, 60473089, 60003006), Liaoning Province Natural Science Foundation of China (Project no. 20032019, 001015), the Scientific Research Foundation for the Returned Overseas Chinese Scholars by SEM of China, Modern Distance Education Engineering Project by MoE of China, and the Research Grants Council of Hong Kong, China (Project no. PolyU 5167/99E).

Table 2 The sorting outcomes

c n	1	2	3	4	5	6	7	8	9	10	11
1	0.00005	0.99955	0.00002	0.00002	0.00009	0.00004	0.00004	0.00004	0.00007	0.00004	0.00004
2	0.17134	0.21221	0.04138	0.03548	0.11356	0.06250	0.06167	0.05687	0.16065	0.05456	0.02978
3	0.92739	0.00700	0.00297	0.00523	0.01089	0.00827	0.01276	0.00425	0.01099	0.00737	0.00288
4	0.17986	0.07280	0.03363	0.04749	0.12518	0.14992	0.05799	0.07223	0.17191	0.06493	0.02406
5	0.05137	0.06989	0.01861	0.02341	0.37800	0.05206	0.03707	0.04958	0.12381	0.04930	0.14690
6	0.00000	0.00001	0.00000	0.00000	0.00001	0.00001	0.00000	0.00001	0.00001	0.00000	0.99995
7	0.00006	0.00002	0.00010	0.99940	0.00007	0.00008	0.00010	0.00005	0.00006	0.00004	0.00002
8	0.00000	0.00000	0.99999	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
9	0.05437	0.03192	0.03682	0.04486	0.13992	0.09657	0.09726	0.20909	0.16097	0.09186	0.03636
10	0.06524	0.03491	0.03010	0.04088	0.19246	0.06786	0.07249	0.10337	0.29883	0.06398	0.02988
11	0.00044	0.00020	0.00014	0.00034	0.00149	0.00130	0.00192	0.99207	0.00135	0.00054	0.00021
12	0.05138	0.01118	0.00669	0.02443	0.06815	0.08666	0.51411	0.13975	0.06400	0.02496	0.00869
13	0.01829	0.00520	0.00469	0.00828	0.02835	0.83461	0.03298	0.02023	0.01643	0.02739	0.00355
14	0.00484	0.00094	0.00080	0.00331	0.00385	0.00901	0.96009	0.01030	0.00339	0.00273	0.00074
15	0.00756	0.00112	0.00094	0.00460	0.00577	0.00744	0.95239	0.00782	0.00631	0.00517	0.00088
16	0.03325	0.01849	0.01734	0.04138	0.11116	0.22466	0.14304	0.09315	0.08211	0.22107	0.01435
17	0.04207	0.01412	0.02615	0.21130	0.10464	0.08048	0.15135	0.12206	0.10669	0.13044	0.01070
18	0.00049	0.00015	0.00013	0.00021	0.00154	0.00101	0.00047	0.00040	0.00084	0.99461	0.00015
19	0.03260	0.01277	0.01440	0.04253	0.16578	0.31176	0.07046	0.06851	0.19229	0.07535	0.01355
20	0.00882	0.00265	0.00360	0.00901	0.02095	0.88637	0.01677	0.01733	0.01607	0.01490	0.00353
21	0.00429	0.00187	0.00093	0.00160	0.93912	0.00605	0.00312	0.00542	0.02548	0.01083	0.00129
22	0.87663	0.00266	0.00277	0.00546	0.02311	0.01662	0.02030	0.01197	0.02759	0.01057	0.00232
23	0.00442	0.00127	0.00135	0.00132	0.02344	0.00356	0.00281	0.00461	0.95140	0.00400	0.00182
24	0.03006	0.00878	0.00927	0.01551	0.36882	0.18518	0.04435	0.08095	0.18081	0.05793	0.01834

REFERENCES

- [1] Schneider G.P. and Perry J.T., *Electronic Commerce*, Cambridge, Mass: Course Technology, 2000.
- [2] Kalakota R., *Electronic Commerce: a Manager's Guide*, Mass.: Addison-Wesley, 1997.
- [3] Hallikasa, J., Karvonenb, I., Pulkkinenb, U., Virolainenc, V.-M. and Tuominen, M., "Risk management processes in supplier networks", *International Journal of Production Economics*, Vol. 90, No. 1, pp47-58, 2004.
- [4] Gunn J., "Extended enterprise integration", *BT Technology Journal*, Vol. 18, No. 2, pp93-99, 2000.
- [5] Tinham B., "Switch on to global integrated business and manufacturing", *Manufacturing Computer Solutions*, Vol. 6, No. 9, pp46-47, 2000.
- [6] Grefen P., Aberer K., Hoffner Y. and Ludwig H., "Crossflow: cross-organizational workflow management

in dynamic virtual enterprises", *Computer Systems Science and Engineering*, Vol. 15, No. 5, pp277-290, 2000.

- [7] Bier, V.M., Haimes, Y.Y., Lambert, J.H., Matalas, N.C. and Zimmerman, R., "Survey of approaches for assessing and managing the risk of extremes", *Risk Analysis*, Vol. 19, No. 1, pp 83-94, 1999.
- [8] Kliem, R.L. and Ludin, I.S., *Reducing Project Risk*, England: Gower Publishing Limited Gower House, 1997.
- [9] Haimes, Y. Y., *Risk Modeling, Assessment, and Management*, 2nd ed, New York: Wiley, 2004.
- [10] Bezdek J C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum, 1981.
- [11] Huang H.Z., *Fuzzy Design*, Beijing: Machinery Industry Press, 1999.

The Research of Product Graphical Information Sharing Technology of Virtual Manufacturing Enterprise in E-Commerce Environment

Mingxu Ma ¹², Haiyue Li ¹², Wenlei Zhang ¹², Chaowan Yin ¹, Chengen Wang ¹

¹ Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

² Graduate School of the Chinese Academy of Sciences, Beijing 100086, China

{ Mingxu MA, Haiyue LI, Wenlei ZHANG , Chaowan YIN, Chengen WANG }@sia.cn

ABSTRACT

This paper has built a product model by UML and corresponding Product Schema. Then we have illuminated transmit mechanism of the product information by a dumbbell XML document. At last, we have pointed out the direction of the research. This research will provide a significative explore to the product data interchange between the members of virtual manufacturing enterprise in e-commerce environment.

Keywords: E-commerce, Product model, UML, XML

1. INTRODUCTION

Along with the rapid development of network technology, the e-commerce technology becomes increasingly ripe. The importance of the e-commerce has been recognized more and more and this kind of trade form has been accepted. Virtual manufacturing enterprise as the important user of e-commerce, comparing with the traditional manufacturing enterprise, has many merits such as sharing infrastructure, outstanding manpower, advanced technology resources, sharing the market risk and researching development cost, sharing market and customer etc. The product data exchange certainly must be carried among the members of virtual manufacturing enterprises in order to share product information. In collaborative design, the visualization technology is one of tools to realize the product data exchange, and can realize the some functions, such as the CAD's documents browsing, marking, reading etc.

At present, many visualization systems (Such as the wiseVIEW of SAMSUNG(SAMSUNG, 2004), the VOLO VIEW EXPRESS of AUTODESK (AUTODESK, 2003), The Browsing Tools of Tsinghua Ying Tai(Ying Tai,2004) etc.) all exist a common question to browse CAD's documents, which is that the documents can be opened by using the corresponding tool only when the documents must have been completely transmitted during document delivery. The more complicated the product increasingly become, the more corresponding huge the product data are also, once completely download will fairly cost time. Sometimes, we needn't to see whole details. When one small part of details of product has been modified, the whole product graphical documents needs to be transmitted again. So the limited network resource is wasted seriously.

2. PRODUCT MODEL WITH UML

Product has many features, such as geometry feature, material feature, manufacturing feature, assembly

feature, etc. The each feature has a corresponding class. This paper has built a product model with UML. This model has represented the detailed model structure, especially the geometry feature of product. Shown as the figure 1, the product class is the highest rank class in the model, has its own Product_ID attribute and Product_Name attribute. The material class, geometry class, manufacturing class and assembly class are the subclasses of product class. The product geometry feature is core feature and other features, it is very important to comprehensively and felicitously represent the product geometry feature..

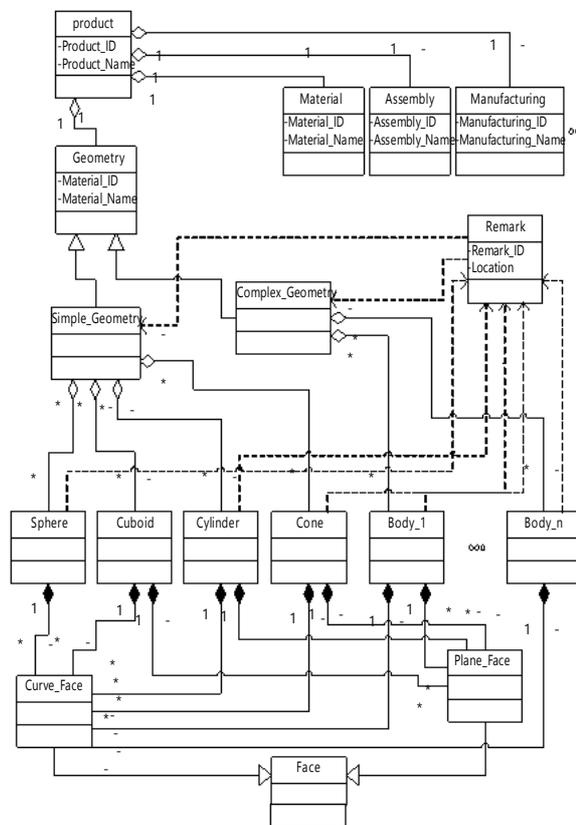


Fig.1 The Product Model

3. THE REPRESENTATION OF PRODUCT DOCUMENT WITH THE SCHEMA OF XML

Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML. Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. XML has many merits, such as extendability, simpleness and ripeness. So this paper has put forward to a modeling method of the product graphical information based on XML to the virtual enterprises in the e-commerce environment.

Shown as the figure3, according to the Schema of XML(Schema,2003), the paper has described a product data model, including product geometry information and its remark information. Owing to the characteristic of XML itself, this model possesses very strong extensibility and still can add other product information except mentioned above the geometry information and the remark information, like the material information, the assembly information, the process information and manufacturing information etc.

```
<?xml version="1.0" encoding="utf-8" ?>

<xs:schema
targetNamespace="http://localhost:8080/Product.xsd"
xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns="http://localhost:8080/Product.xsd">
  <xs:annotation>
    <xs:documentation>
      This is a product model written by XML
      Schema.
    </xs:documentation>
  </xs:annotation>
  <xs:element name="product_model">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="Geometry"
minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="Material"
minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="Assembly"
minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="Manufacturing"
minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="Remark"
minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
      <xs:attribute name="product_ID"
type="xs:ID" use="required"/>
      <xs:attribute name="product_Name"
type="xs:string" use="required"/>
    </xs:complexType>
  </xs:element>
```

```
<xs:element name="Material"/>
<xs:element name="Assembly"/>
<xs:element name="Manufacturing"/>
<!-- esle product feature -->
<xs:element name="Geometry">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="Simple_Geometry"/>
      <xs:element ref="Complex_Geometry"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="Simple_Geometry">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="Sphere"
minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="Cuboid"
minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="Cylinder"
minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="Cone" minOccurs="0"
maxOccurs="unbounded"/>
    </xs:sequence>
    <!-- esle Geometry feature -->
  </xs:complexType>
  <xs:attribute name="ID" type="xs:ID"/>
</xs:element>
<xs:element name="Complex_Geometry">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="Entity" minOccurs="0"
maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:attribute name="ID" type="xs:ID"/>
  </xs:complexType>
</xs:element>
<xs:element name="Remark">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="Title"/>
      <xs:element ref="Content"/>
      <xs:element ref="Location"/>
      <xs:element
ref="Correlative_entity_ID"/>
    </xs:sequence>
    <!-- esle Remark condition -->
  </xs:complexType>
  <xs:attribute name="remark_ID"
type="xs:ID" use="required"/>
</xs:element>
```

```

<xs:element name="Title" type="xs:string"/>
<xs:element name="Content" type="xs:string"/>
<xs:element name="Location"
type="relative_coordinate"/>
<xs:element name="Correlative_entity_ID"/>
<xs:element name="Sphere" type="SphereType"/>
<xs:complexType name="SphereType">
  <xs:all>
    <xs:element ref="Sphere_center"/>
    <xs:element ref="Radius"/>
    <xs:element ref="Color"/>
  </xs:all>
  <xs:attribute name="ID" type="xs:ID"/>
</xs:complexType>
<xs:element name="Sphere_center"
type="relative_coordinate"/>
<xs:element name="Radius" type="xs:double"/>
<xs:element name="Color" type="xs:string"/>
<xs:element name="Cylinder_bottom_center"
type="relative_coordinate"/>
<xs:element name="High" type="xs:double"/>
<xs:element name="Cylinder">
  <xs:complexType>
    <xs:sequence>
      <xs:element
ref="Cylinder_bottom_center"/>
      <xs:element ref="Radius"/>
      <xs:element ref="High"/>
      <xs:element ref="Color"/>
    </xs:sequence>
    <xs:attribute name="ID" type="xs:ID"/>
  </xs:complexType>
</xs:element>
<xs:element name="Cuboid"/>
<xs:element name="Cone"/>
<xs:element name="Entity"/>
<xs:complexType name="relative_coordinate">
  <xs:sequence>
    <xs:element ref="x"/>
    <xs:element ref="y"/>
    <xs:element ref="z"/>
  </xs:sequence>
</xs:complexType>
<xs:element name="x" type="xs:double"/>
<xs:element name="y" type="xs:double"/>
<xs:element name="z" type="xs:double"/>
</xs:schema>

```

Fig.2 The Product Schema

4. THE EXCHANGE OF PRODUCT INFORMATION

XFI(XFI,2002) allows the exchange of specific information of interest without sending complete XML documents. The most important merit of XFI is to reduce the quantity of communication data and save cost of communication. According to the product Schema having been created in the second paragraph, we have built a simple dumbbell example to show the application of XFI in the data transmission. Then, we have defined the mechanism to create and transmit the product graphical documents and their attached remark. The Figure.3 is a example of a product XML document to describe a dumbbell.

```

<?xml version="1.0" encoding="utf-8"?>
<product_model
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://localhost:8080
Product.xsd"
xmlns="http://localhost:8080/Product.xsd"
product_ID="P00001">
  <Geometry>
    <Simple_Geometry ID="S00001">
      <Sphere ID="Sp00001">
        <Sphere_center>
          <x>0</x>
          <y>0</y>
          <z>0</z>
        </Sphere_center>
        <Radius>1.0</Radius>
        <Color>red</Color>
      </Sphere>
      <Sphere ID="Sp00002">
        <Sphere_center>
          <x>4.0</x>
          <y>0.0</y>
          <z>0.0</z>
        </Sphere_center>
        <Radius>1.0</Radius>
        <Color>blue</Color>
      </Sphere>
      <Cylinder ID="C00001">
        <Cylinder_bottom_center>
          <x>0.0</x>
          <y>0.0</y>
          <z>0.0</z>
        </Cylinder_bottom_center>
        <Radius>1.0</Radius>
        <High>4.0</High>
        <Color>black</Color>
      </Cylinder>
    </Simple_Geometry>

```

```

    <Complex_Geometry/>
  </Geometry>
  <Remark remark_ID="R00001">
    <Title>Geometry Remark</Title>
    <Content>the radius is too thick!</Content>
    <Location>
      <x>2.0</x>
      <y>2.0</y>
      <z>2.0</z>
    </Location>

    <Correlative_entity_ID>P00001</Correlative_entity_ID>
  </Remark>
</product_model>

```

Fig.3 The dumbbell XML document according to the Product Schema

4.1 The validity of the fragment in a product XML document

The fragment in a product XML document is a general term to refer to part of an XML document, plus possibly some extra information, that may be useful to use and interchange in the absence of the rest of the XML document. We can appoint the context information of a product graph documents fragment by transmitting the part information rather than the whole product graph documents. So we can achieve the aim to reduce the data transmission quantity.

A valid product XML fragment has a continuous test block according to XML specifications. The figure 4 shows some valid fragments.

- a. 0.5
- b. <Radius>0.5</Radius>
- c. <Content>the radius is too thick!</Content>
- d. <Location>


```

        <x>2.0</x>
        <y>2.0</y>
        <z>2.0</z>
      </Location>

```

Fig.4 Some valid product XML fragments

4.2 The transmit mechanism of the product XML fragment

By the most important merit of XFI, we can build a mechanism to realize the aims to reduce the quantity of communication data .

For example, we want to change the radius of the dumbbell' handle , we shouldn't need to transmit whole

dumbbell model data before getting to the part of interest. The goal of this paragraph is to define a way to enable processing of small parts of an product XML document without having to process everything up to the part in question. This can be done regardless of whether the parts are entities or not, and the parts can either be viewed immediately or accumulated for later use, assembly, or other processing.

The following content is the XML fragment document with the right fragment about the change of the dumbbell handle' radius.

```
<Radius>0.5</Radius>
```

And its context following shows as the figure 5:

```

<?xml version="1.0" encoding="UTF-8"?>
<f:fcs
xmlns:f="http://www.w3.org/2001/XMLSchema-fragment"
parentref="http://localhost:8080/Product.xml"/>
<product_model>
  <Geometry>
    <Simple_Geometry>
      <Cylinder>
        <f:fragbody
fragbodyref="http://localhost:8080/dumbbell_radius.xml"/>
      </Cylinder>
    </Simple_Geometry>
  </Geometry>
</product_model>

```

Fig.5 The context of XML fragment about the radius of the dumbbell' handle

The context is information to describe the location of the original document fragment. The context include the URL of the original document, the father information of the fragment body, etc. For transmit the fragment information, we can transmit the fragment document and its context to the receiver. The receiver can insert the fragment to the correct location in the original document by the context. In this way, we have realized the transmit mechanism of the product XML fragment.

5. CONCLUSIONS AND FURTHER WORK

This paper has built a product model by UML and corresponding Product Schema. Then we have illuminated transmit mechanism of the product information by a dumbbell XML document. To realize the visualization mechanism, we want to use the Java3D technology. Java3D is a low level 3D scene-graph based graphics programming API for the java language.

Each of the members of virtual manufacturing enterprises in e-commerce can change CAD's documents of oneself into the product graphical document that this paper has built, and show the product graphical document by using the corresponding browsing tool, then achieve the purpose of low cost information sharing owing to the open character of XML. This research will provide a significative explore to the product data interchange between the members of virtual manufacturing enterprise in e-commerce environment.

ACKNOWLEDGEMENT

The research is supported by the Chinese Hi-Tech R&D

program under the Grant (2002AA414420)

REFERENCES

- [1] WiseVIEW of SAMSUNG. <http://www.sds.samsung.com/product/wiseview/wiseview.html>. 2004.
- [2] VOLO VIEW EXPRESS of AUTODESK. <http://usa.autodesk.com/>.2003.
- [3] Tsinghua Ying Tai. <http://www.thit.com.cn/>.2004.
- [4] Schema of XML. <http://www.w3.org/XML/Schema>. 2003-4-29.
- [5] XFI. <http://www.w3.org/TR/xml-fragment>. 2002-3-29.

Virtual Enterprises Risk Management DSS under Electronic Commerce

Min Huang¹, Hongmei Yang¹, Xingwei Wang¹, W.H.Ip², K.L. Yung²

¹ Box 135, Faculty of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, China, mhuang@mail.neu.edu.cn

² Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China

ABSTRACT

Risk management is important to the development of enterprise as well as social-economic prosperity. Virtual enterprise is the potential mode of future enterprise under electronic commerce environment and the risk management for it is a popular research area recently. Due to the complexity of its risk management a decision support system (DSS) with 3-bases-1-cell structure was designed. By coordinating data base, model base, algorithm base and dialogue cell, the functions of project management, risk identification, risk assessment, risk evaluation and risk programming was supported. The user-friendly system has such main characteristics as generality for verity virtual enterprise as well as different projects and the flexibility of model and algorithm, ensuring a standardized, scientific and informational risk management for virtual enterprises.

Keywords: virtual enterprise, risk management, decision support system, electronic commerce

1. INTRODUCTION

The subject of risk management has a history dating back to the 1950s. In the 70s it rapidly developed theoretically and practically and prevailed among developed countries [1-4]. With the formation of globalized market economy and emerging of electronic commerce, Virtual Enterprise (VE) has become a potential mode of enterprise development. Such a complicated environment requires effective production management as well as a complete and practical Decision Support System (DSS) for risk management to assist decision makers in making objective and correct decision.

Risk management is significant to the development of enterprise as well as social-economic prosperity. It is generally defined as an enterprise management which includes identification and assessment of every potential risk, as well as proper manipulation and solution, aiming at ensuring a safe enterprise operation with the lowest risk cost [5,6]. Trouble-free risk management relies on scientific strategies and its process consists of risk identification, risk assessment and risk evaluation sequentially and response-measure programming based on the first 3 steps so as to carry out effective risk manipulation and solution to unfavorable consequences. VEs are temporal virtual organizations for grasping some newly occurred opportunity [7,8]. It is always organized by project mode. The projects are different from each other for different VE. So, the risk management should be different accordingly. DSS of VE provides a functional support to risk identification, risk assessment, risk evaluation and risk programming for different projects in order to ensure a safe and effective operation for enterprises.

2. LOGICAL DESIGN FOR SYSTEMATIC STRUCTURE

Characteristically, project risk models differ when VEs are facing different opportunities, which require a flexible risk-describing function of the system. Similarly, the complexity of problems in a projects calls for flexible algorithmic solutions. Based on such requirements, a 3-base-1-cell logical structure [9] is established (see Figure 1), which consists a database, a model base, a algorithm base and a dialogue cell.

The logical structure of the system effectively carries out the 4 operational functions (risk identification, risk assessment, risk evaluation and risk programming) by constructing an operational environment of DSS for VE in the dialogue system on the informative basis of the 3 bases (database, model base and algorithm base) with the assistance of friendly man-machine interface and dialogues.

(1) Dialogue System

This is the component that users work with directly, the general frame of which is constructed according to the basic flow process of risk management. From a user's point of view, the frame presents itself as a master-control menu.

(2) Model Base

1) Requirements of VE DSS to model base

Risk evaluation model is a multi-level synthetic evaluation model with a dendritic structure (see Figure 2).

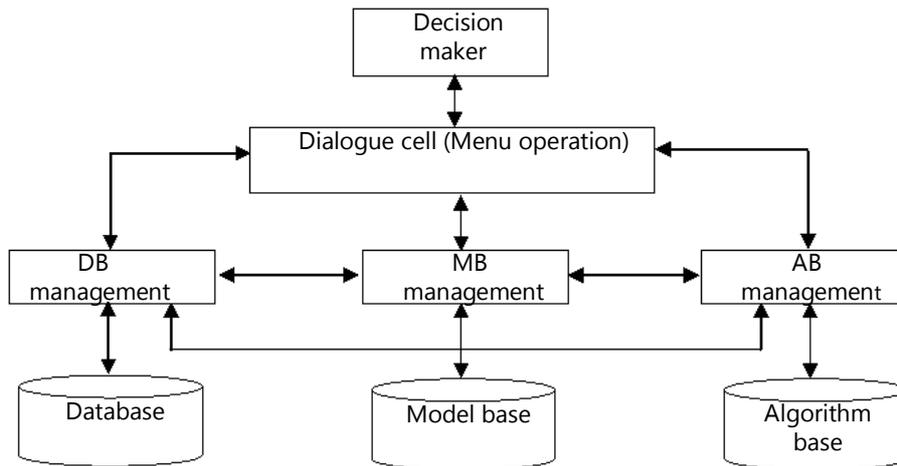


Figure1 The logic structure chart of the system

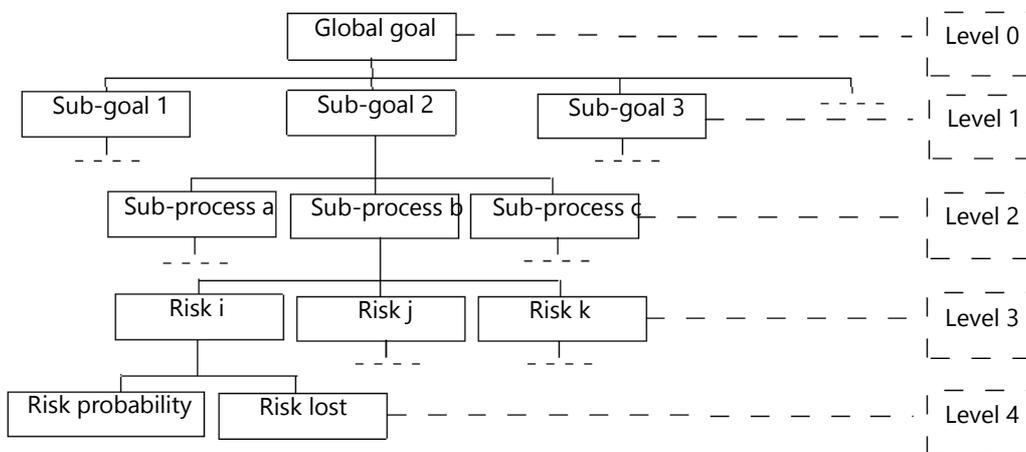


Figure 2 The hierarchic model for risk evaluation

When applied to risk management of different VE, the avoid the inconvenience of repetitive maintenances. model has the same levels but different numbers of node. Therefore a versatile structure of database is established according to the characteristics of the system. Consequently, managing system of model base should be able to establish a special model according to specific requirements of a project, which can be realized by achieving the followings:

ÇSeparating model from algorithm

Model must be separated from algorithm; otherwise the source program has to be modified so as to make algorithm adaptive to a changed model, which is impossible for the operators to do by themselves.

ÇAutomatic model generation

Unlike some DSS, system programmers cannot store some fixed models of DSS for VE in computers because the structure of the model is determined by a specific VE. Hence, it is necessary to build up a versatile system to assist users in establishing and storing a model structure according to his business.

ÉDatabase structure with adaptability to all models

Such a database structure should be established so as to modify itself, complying with a changing model base and

2) Model base design

In this part, the core is how to store a model structure in computers. This research expresses the structure in a 3-level frame, which combines frame-knowledge and characteristics of database system and fully presents the external and internal features of the model, presenting a simple and straightforward visual illustration and proving the easiness of establishment.

In Frame level 1 store the model name, date of establishment and referential information etc., which are all external characteristics of the model.

Frame level 2 describes the internal characteristics of the model, mainly including the number of levels and nodes. Since 5 fixed levels are stored in the system, the number of nodes determines the main feature. The content consists of the name of model, level names, and the number of nodes in each level, etc.

Frame level 3 presents the physical mode of the model evaluation and risk programming. The foreign-key reference relations are shown in Figure 3.

Different model has different internal physical mode. Based on the characteristic that the synthetic evaluation reflects on each nodal point in this level, its content includes model name, level name, node name and the weight of the node etc.

(3) Database

Database is an important component of DSS. In this system, 9 tables are designed based on the data of risk

(4) Algorithm Base

This system provides the risk-programming model with two methods: genetic algorithm and Tabu search, which are stored in database system for users' options. Further the new algorithm can be added, and the exist algorithm can be deleted and modified by user according to practical needs, providing flexibility in algorithms.

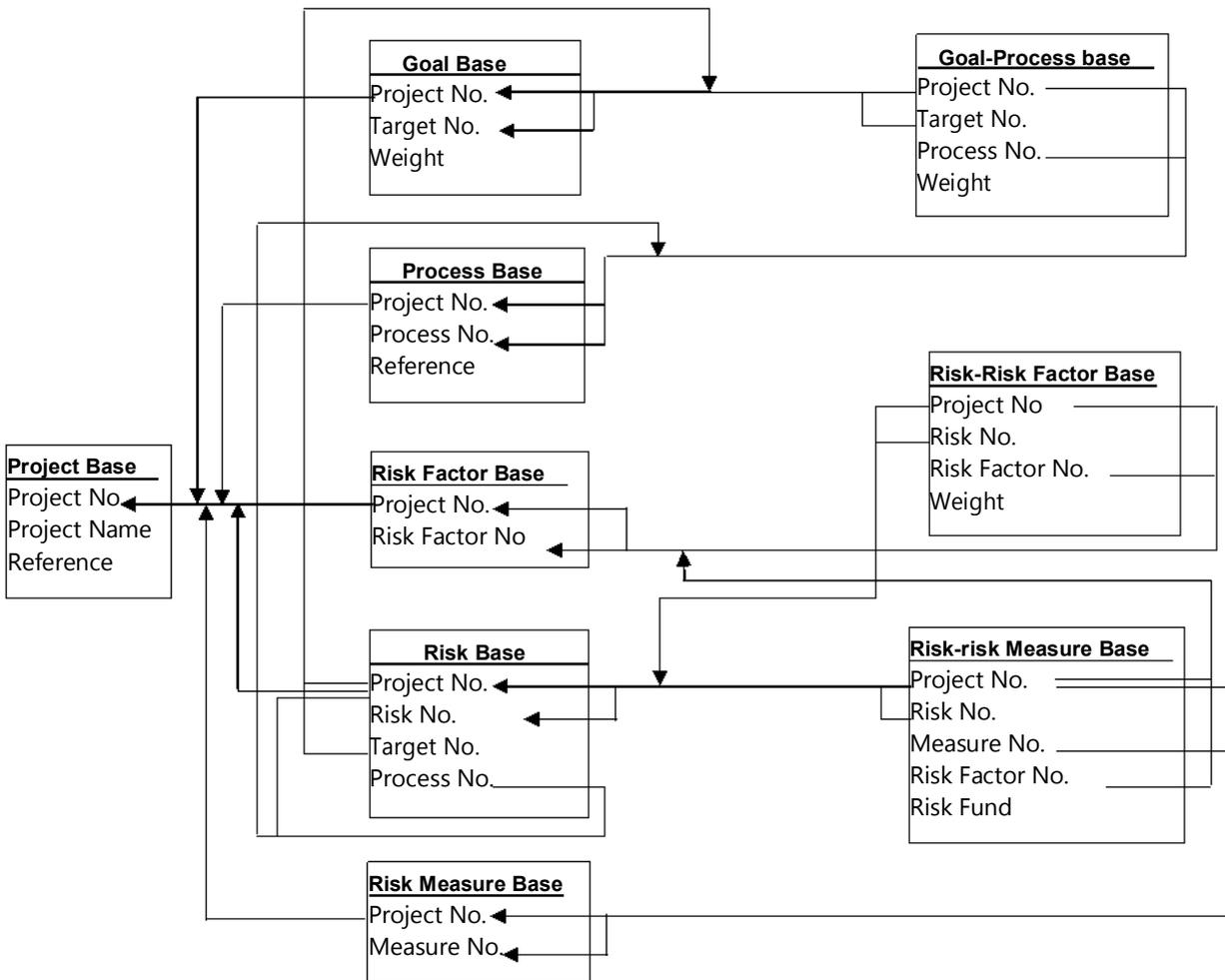


Figure 3 Relationship chart of foreign key

3. SYSTEM-FUNCTIONAL MODULES

In accordance with the demands of VE risk management, this decision support system can be functionally divided into 5 parts as follows:

(1) Project Management is an all-project-oriented management of VE, providing such projective functions as building, opening and deleting projects.

(2) Risk Identification is the first step of risk management. According to risk evaluation model, the managed content of this stage is the global goal, sub-goal, sub-processes, risks, risk localization, risk manipulation method etc. It also provides inputting, searching, sorting, inserting, modifying and saving functions for the data. This function corresponds with the establishment of frame level 1 and 2.

(3) Risk Assessment is a process in which the identified

risks are quantitatively estimated, including sub-goal weight assessment, sub-goal and sub-process relevant level assessment, relevant risk weight assessment, risk factor weight assessment and fuzzy description to pre-manipulation and after-manipulation risk factors. The module also provides inputting, searching, sorting, modifying and saving functions for the data. This function corresponds with the establishment of frame level 3.

- (4) Risk Evaluation calculates the general risk level according to risk evaluation model (Figure 2), where the risk evaluation is dealt from local (lowest level) to global system (level 0).
- (5) Risk Programming provides the optimized risk-manipulating strategies according to programming model. Such function is especially applied to algorithm base management and algorithmic operation.

4. CONCLUSION

In this research, decision support system for VE under electronic commerce is designed. This user-friendly system has such main characteristics as generality for verity VE as well as different projects and the flexibility of model and algorithm. The system driven by general strategic goal of enterprise, assist the decision maker in establishing a standardized, scientific and informational risk management in an electronic commerce environment full of internal and external uncertainties.

ACKNOWLEDGEMENTS

The authors wish to thank the support of the National Natural Science Foundation of China (Project no.

70101006, 60473089, 60003006), Liaoning Province Natural Science Foundation of China (Project no. 20032019, 001015), the Scientific Research Foundation for the Returned Overseas Chinese Scholars by SEM of China, Modern Distance Education Engineering Project by MoE of China, and the Research Grants Council of Hong Kong, China (Project no. PolyU 5167/99E).

REFERENCES

- [1] Ralph L K and Irwin S L., *Reducing Project Risk*, Aldershot: Gower Publishing Limited Gower House, 1997.
- [2] Bier, V.M., Haimes, Y.Y., Lambert, J.H., Matalas, N.C. and Zimmerman, R., "Survey of approaches for assessing and managing the risk of extremes", *Risk Analysis*, Vol. 19, No. 1, pp 83-94, 1999.
- [3] Haimes, Y. Y., *Risk Modeling, Assessment, and Management, 2nd ed*, New York: Wiley, 2004.
- [4] Lu Y.J., Lu J.Y., *Project risk Management*, Beijing: Tsinghua University Press, 1998.
- [5] Zheng C.T., *Risk Management: Theory and Application*, Taipei: Wu Nan Press, 1995.
- [6] Mars G. and Weir D., *Risk Management*, Aldershot: Ashgate Publishing Limited, 2000.
- [7] Park K.H. and Favrel J., "Virtual enterprise - information system and networking solution", *Computers and Industrial Engineering*, Vol. 37, No. 1, pp441-444, 1999.
- [8] Ouzounis V. and Tschammer V., "Framework for virtual enterprise support services", *Proceedings of the Hawaii International Conference on System Sciences 1999*, pp297-305, Los Alamitos, Jan.3-6, 1999.
- [9] Gao H.X., *Decision Support System: Theory, Method and Case*, Beijing: Tsinghua University Press, 2000.

Virtual Manufacturing and Virtual Spaces

Mélissa Saadoun

INEDIT Institute, 24 rue du Panorama - 95800 Cergy St-Christophe, France
msaadoun@aol.com

ABSTRACT

This article presents news problems linked to the increasing use of **virtual spaces** in the building of new types of manufacturing what is currently called **Virtual Manufacturing (VM)**. VM has many consequences in the activity of industrial companies whatever be size, economic sector or geographical area. Particular attention is paid on virtual spaces, manufacturing and collaboration, skills matrix and manufacturability, the CAD and Virtual Prototyping case and, finally, the virtual partnerships or enterprise as a new type of organization.

Keywords: Virtual Enterprise, Virtual Manufacturing, Virtual Partnership, Virtual Prototyping, Virtual Space.

1. INTRODUCTION

Virtual Manufacturing (VM) can be defined as an integrated, synthetic manufacturing environment exercised to enhance all levels of decision and control. The vision of VM is to provide a capability to «Manufacture in the Computer ». That means, VM will provide a modeling and simulation environment so powerful that the fabrication/assembly of any product, including the associated manufacturing processes, can be simulated in the computer.

It is then a manufacturing within a computer supported on the network, having a strong links with real or physical world of manufacturing (See Figure 1).

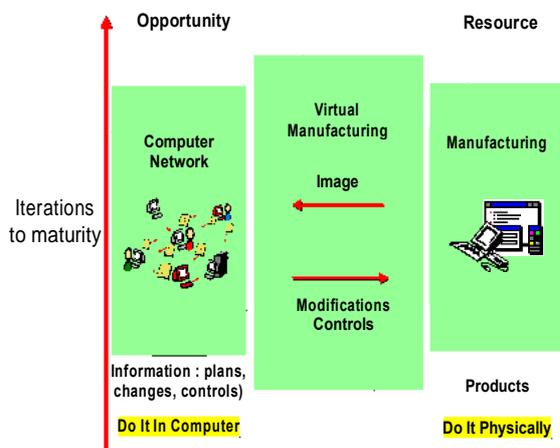


Fig.1. Virtual Manufacturing Vision

The main significant (semantics) of this definition are:

- *Synthetic or virtual working space*, mixing real and simulated objects, activities and environments of processes which supports the construction and use of distributed manufacturing simulations by synergistically providing a collection of analysis tools, simulation tools, implementation tools, control tools, models (product, process and resource), equipment, methodologies and organizational principles (culture) exercising , constructing and executing specific manufacturing

simulations using the environment.

- *Enhance*: increase the value, accuracy, validity levels: from product concept to disposal, from the shop floor to the executive suite, from factory equipment to the enterprise and beyond, from material transformation to knowledge transformation decision: understand the impact of change (visualize, organize, identify alternatives).
- *Control*: predictions effect actuality.

2. IMPACTS OF VM AND NEEDS

VM affect the all industrial company activities starting from the manner in which market signal are taken into account, its relations with customers and providers until its internal reorganization or reconfiguration. In particular here are concerned the impact on the company memory, capital investment, design of production; cost estimation, risk management, client interfaces, functionality interfaces, workshop. The needs for VM are:

- Technology and development for models and simulations.
- Virtual infrastructures.
- Tools for working in the virtual spaces such as, the CAD kernel.

The scope of VM ranges from workshop to company areas. To set up a theoretical approach of this phenomenon, we propose two cases : CAD and virtual prototyping. Both will evolve more in the area of the future workshop define under this : from the «virtualisation of the real to the realization of the virtual », characterizing the new productive reality.

3. SHARED SPACES AND COLLABORATION

One of the main problems with classical forms of collaboration: they are not able to capture the knowledge generated during the collaboration process. In reality, most of the collaboration process are ad-hoc process addressed to search solutions to different emerging problems but without an infrastructure that support this process. In that case the available knowledge, generated by collaborations, is used only one time and, probably,

lost for ever [1].

We must create simple models to capture this knowledge and avoid dissipating the expertise created by collaboration [2]. Collaboration is essentially based on knowledge interaction between complementary expertise, but for becoming reality there should be a common understanding allowing creating new concepts. One of these new concepts is that of the virtual spaces. Virtual spaces have a particular nature in which problems such as communication coordination or synchronisation are among the most relevant [3].

The distributed and shared infrastructure is composed of virtual spaces constituted of virtual reality models that could be manipulated by complementary expertise's of trading partners. Those virtual creations and manipulations by multidisciplinary teams enable people to reach a common understanding and to elaborate shared concepts in a better and faster way than any long discussion during meetings. Virtual spaces shared by disciplines provide a collective sense of identity and fundamental purpose as requested by the knowledge-creating company. One practical problem is to identify the rationale of any interaction need and interdependencies between entities manipulated by complementary disciplines and, to be able to manage the consistency of generated constraints. The virtual project office example of virtual shared space will be demonstrated in the, where several industrial trading partners forming a consortium will use the virtual project office as a distributed infrastructure (see Figure 2).



Fig.2. Example of virtual shared space : the virtual project office (Source EPICE ESPRIT project)

Expected virtual project office services to be distributed to the trading partners include the following:

- Work breakdown structure (WBS) and organization breakdown structure (OBS) management on the Web.
- Workflow management and tasks coordination.
- Requirements and constraints consistency management.
- Project progress monitoring.
- Configuration and data management.

4. VM INFORMATION BY INTERNET: THE CASE OF CAD DATA

HTML documents can be useful to send some kind VM information, transmitting design and manufacturability data to Web navigators from remote sites. Examples of this include STEP files, CAD models and textual data such as results of design or fabricability or information based on manufacturing catalog. One limit to a fluent data transmission is the large bandwidth at a given moment. For example, if we want to have a view of CAD image with a Web navigator, we need several minutes of transmission and interpretation. So the interactive operations (such as rotation of one CAD model) become impossible in a real time.

For some types of VM data, it is possible to get a larger bandwidth for transmitting data as bitmapped images but using VRML (Virtual Reality Modeling Language). One other solution of the congestion problem is to transmit only the specific code for programs to be executed in local sites. Any code must be involved but it is better to use a code written in a specific language for that purposes such as HotJava or Telescript. They are written for exchanging software components over the network. At the arrival site, it run programs and enable users to interact each other. Using these generic tools, one can compose interactive programs containing several types of documents. The navigator HotJava permit to interact dynamically without congesting the network with additional traffic. To do it, HotJava adds a new type of markup, the HTML APP, where APP means «applet», a program wrote in Java language and running using HotJava which can link dynamically a Java code from the host and execute on the local machine.

Linked to these tools, there are the development environments of applications. These tools are used to build packaged software modules «add on» which will be coupled with other tools. In CAD soft market, examples include Pro/Develop from Pro/Engineer, GRIP from EDS/Unigraphics, the Development Language (MDL) of Bentley Systems. They can be used to create software modules of network to share, execute or buy on the network starting from a simple CAD environment.

On the implementation side (Figure 3), the Internet and Web-related technologies are used through Intranet (Internet tools used on an enterprise private network) and Extranet (Internet tools used on a trading partners private network). Otherwise, encryption systems and large bandwidth on the Internet must allow trading partners to avoid the use of a costly private network. The security aspect should not be underestimated as it was rated as the second greatest challenge in implementing collaboration between trading partners.

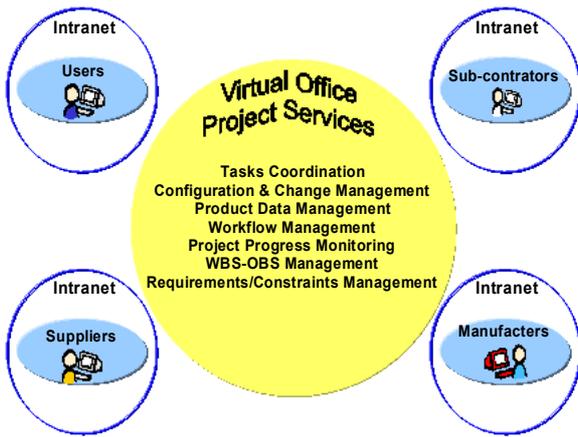


Fig.3. Example of a virtual shared space implementation

5. VM: THE CASE OF VIRTUAL PROTOTYPING

Virtual prototyping models offer an entirely new perspective in particular for the Internet trading partners and provide a strong sense of reality and the means to test the model of a product through meaningful interaction. All-digital prototypes resemble their physical counterparts as closely as possible in terms of visual three-dimensional images, behavioral, haptic, and auditory characteristics.

An Internet virtual salesroom concept is needed to be based on a virtual three-dimensional photo realistic salesroom environment featuring a selection of virtual product models (Figure 4). A particular enabling technology, virtual prototyping, in the context of consumer electronics and telecommunications industries is addressed (in Nokia case). Virtual prototyping is the application of advanced modeling and simulation techniques, user interface techniques and virtual reality techniques to support electronic commerce for customers and also business-to-business electronic commerce in the value chain that is product design, manufacturing, and sales [4].

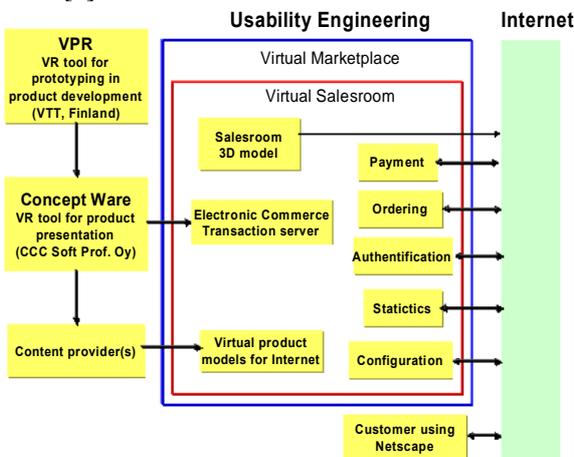


Fig.4. the proposed concept for Internet electronic commerce utilizing digital virtual product models in marketing and sales

Virtual prototyping means the process of using virtual prototypes instead of or in combination with physical prototypes, for innovating, testing and evaluating of specific characteristics of a candidate design. The following aspects are dominant:

- Functionality, created virtually, clearly defined and realistically simulated (for example product functionality and dynamic behavior).
- Human Interaction: the human functions involved must be realistically simulated, or the human must be included in the simulation (that is real-time operator-in-the-loop simulation).
- Environment: an off-line (non-real-time) computer simulation of the functions can be carried out, or a combination of computer (non-real-time) and hardware-in-the-loop (real-time) simulation can be carried out.

In virtual prototyping, a virtual world comprises both the VR model of a product under development (that is virtual prototype), and the model of the product's target environment, called virtual target environment [5]. Virtual worlds, the VR models, are highly advanced human-computer interfaces based on virtual reality techniques, providing a strong intuitive sense of reality and the means to change the world through meaningful interaction. Virtual worlds may include various levels of reality, thus enabling us to concentrate on any desired aspects of the product.

The most elementary interaction interfaces can be a keyboard and a mouse with a conventional window-based two-dimensional user interface. A more sophisticated virtual prototyping environment with a 3D user interface may include, for example a head mounted display, a 3D position and orientation tracking device, and auditory and haptic feedback devices. The type of representation in a virtual prototyping environment may be spatial, *n*-dimensional with abstract information spaces, or various combinations of video, natural and computer-generated images. During the development process, a virtual world can be seen as a model that best matches reality.

A necessary aspect of the full exploitation of the results of virtual prototyping is the requirement for the communication of the results to the interested parties. These may include persons taking part in actual product development - digital signal processing designers, software and hardware designers, and mechanical designers - or people involved in the sales and marketing of the product - marketing and sales representatives, wholesale and retail representatives, and of course customers. Quite often these persons may be quite widely geographically distributed, so solutions based on utilizing communication networks are needed.

The Internet is a viable medium for communicating of the results of virtual prototyping to parties in distributed concurrent product development. Enterprise benefits arise from the potential of the Web as a distribution channel, a

medium for marketing communications, and a market in and of itself [6]. Customer benefits arise primarily from the structural characteristics of the medium and include availability of information, provision of search mechanisms, and on-line product trial, all of which can lead to reduced uncertainty in the purchase decision. It is especially the last one where virtual prototyping would seem a superior technology.

An opportunity arises when the offering is differentiated by elements such as convenience through direct electronic distribution of software, or enjoyment through a visually appealing and exciting Web site. A visually appealing photorealistic operational virtual 3D prototype of a product would certainly fall in the same category. As evidence that this kind of behavior is already happening on the Web, consumers indicated that price was the least important product attribute when making on-line purchases [7].

6. MANUFACTURABILITY AND MATRIX OF COMPETENCIES

Industrial companies must have a virtual presentation of its kernel competencies using, for example, appropriated Internet servers as brokers of these competencies. Each company has their own competency portfolio, develop and enrich it continuously, and supplying then virtually. At this point, Manufacturability Analysis is an essential step in defining VM. This is a measure of the effort required to manufacture the part according to the design specifications (estimate the manufacturing time and cost). Given a computerized representation of the design (i.e. a solid model) and a set of manufacturing resources, the automated manufacturability analysis problem can be defined as follows: Determine whether or not the design attributes (e.g., shape, dimensions, tolerances, surface finishes) can be achieved. Two alternatives appears:

- The design is found to be manufacturable, then determine a manufacturability rating, to reflect the ease (or difficulty) with which the design can be manufactured.
- The design is not manufacturable, then identify the design attributes that pose manufacturability problems.

A VM tool, operating across the Internet, systematically generates redesign suggestions to improve the manufacturability of product designs while still satisfying the design's functional requirements. Linked to the Manufacturability Analysis, there is a Manufacturing Resource Matrix. In this case the system uses a tool for identification of potential participants for supplying manufacturing parts of virtual enterprise. It identifies industrial who are able to manufacture automatically a part and have the corresponding capacity to do it. Finally, we have the Parts Characterization Matrix, an other VM tools, which must operate automatically across the distributed manufacturing network.

7. VIRTUAL PARTNERSHIPS & ENTERPRISES

VM as a powerful capability would take into account all of the variables in the production environment from shop floor processes to enterprise transactions. In other words, VM will accommodate the visualization of interacting production processes, process planning, scheduling, assembly planning, logistics from the line to the enterprise, and related impacting processes such as accounting, purchasing and management.

VM is linked to the Virtual Prototyping, the Virtual CAD and Virtual CAM made most of the time by simulation [8]. But these technical solutions need to be completed by a special attention on people working on. The first problem here is the building of the appropriated engineering teams. The second problem is that of the building of virtual partnerships forming Virtual Enterprises (VE). In both cases the manufacturability and skill matrix are the essential constituents. The constitution of VE, for example, needs an electronic presentation and qualification partners. Here's intervene virtual spaces because all this process exists only in that space.

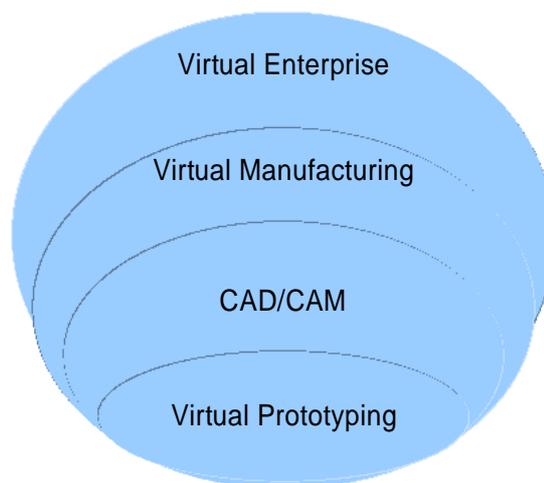


Fig.5. VM related notions

But an other double challenge is appearing: to use some techniques of virtual reality to build, evaluate and test VE, and to combine it with the ad hoc techniques developed within the framework of VE building such as, for example, the VE algebras that we can produce according to every project. We think VE exist only by project [3]. New research problems emerges, one of them being that considering communities as a virtual reality entities. The separated fields must be then integrated, creating a virtual lifecycle. Figure 5 illustrates de relations between all these notions.

PERSPECTIVES

In the short term, the most promising fields for applying VM by Internet can fill the next criteria: to be operational on Internet, every user interaction with a remote server must respect this rule: do not use the tool frequently and

do not need of a large bandwidth. It is better to have a tool serving a large group of users because this eliminate the need for doing copies of tools to the local site users. Having this criteria in mind, we can propose as potential candidate for VM by Internet the data CAD translation, the distributed manufacturing and the design of production systems. Starting from these cases, we can build different levels of scenarios for using VM techniques by Internet.

REFERENCES

- [1] M. Saadoun, Technologies de l'information et management, Hermès, 2000.
- [2] M. Schrage, No More Teams ! Mastering the Dynamics of Creative Collaboration, Doubleday, 1995.
- [3] M. Saadoun, Le Projet Groupware, Eyrolles, 1996.
- [4] M. Sarker, B. Butler & C. Steinfeld, Intermediaries and Cybermediaries: A Continuing Role for Mediating Players in the Electronic Marketplace, Journal of Computer-Mediated Communication, December, 1995.
- [5] P. Svasta, I. Cristea & N. D. Codreanu, Virtual Proptotyping for low-cost electronic packaging, In Proceedings First International Symposium on Concurrent Enterprising Romania, ESoCE'98 4-6 June 1998.
- [6] D. Hoffman, T. Novack & P. Chatterjee, Commercial Scenarios for the Web: Opportunities and Challenges, Journal of Computer-Mediated Communication, December, 1995.
- [7] S. Gupta, A Research Project on the Commercial Uses of the Web, Hermès, 1995.
- [8] B. Stefanescu, Coifa & H. S. Elian, Virtual Reality Systems with applications in the reconstruction of historical monuments, In Proceedings First International Symposium on Concurrent Enterprising Romania, ESoCE'98 4-6 June 1998.

Web Services –An Enabling Technology for Trading Partners Community Virtual Integration

Siew poh Lee, Hanboon Lee, Engwah Lee

Singapore Institute of Manufacturing Technology, 71, Nanyang Drive, Singapore 638075,

Tel: +65-67938383, Fax: +65-67916377

{splee, hblee, ewlee}@SIMTech.a-star.edu.sg

ABSTRACT

The Application Service Provider (ASP) model is a widely adopted business model for communities of buyers and sellers to conduct commerce over the Internet in a virtually integrated environment. However, there are limitations imposed by the ASP model. This paper discusses how an alternative combination of peer-to-peer and web services technologies can provide a suitable, low-cost IT infrastructure that could potentially overcome these limitations and enable trading partners in a community to achieve virtual integration. Such an alternative infrastructure would allow trading partners to improve upon their resource utilisation for IT and, instead, focus on their core businesses, discover new business opportunities and form new partnerships dynamically as and when needed, in a loosely-coupled manner.

Keywords: collaborative commerce, distributed trusted network community, peer-to-peer, virtual integration, web services.

1. INTRODUCTION

The current economy faces two key challenges, namely the market challenges and catching up with the rapid advancement of information and communication technologies (ICT). The market challenges are attributed to factors such as the shift from a PUSH¹ to a PULL² economy, shorter time-to-market, erosion of product loyalty, internationalisation and globalisation, and decentralisation of manufacturing activities [2][3][4]. The other great challenge relates to making sense of rapidly evolving ICT, attributed to the fusion of computer and communication technologies, the advent of the Internet and the wide adoption of open and de-facto standards, such as XML, Java™, TCP/IP and HTTP.

Due to these challenges, businesses from various industry sectors have gradually migrated away from their traditional way of conducting business such as the use of the telephone, fax machine, and postal mail, to conducting their transactions over the Internet. For instance, trading partners in the high-tech manufacturing industries are moving away from the rigid, heavyweight, and tightly-coupled EDI interfaces to the XML-based RosettaNet® [10] business-to-business (B2B) standard, to conduct e-business over the Internet.

2. THE RISE OF THE NEW DIGITAL ECONOMY

Technology has revolutionised the way of doing business

¹ In a PUSH economy, there is more demand than supply. The producer or manufacturer pushes out their products to the consumer.

² In a PULL economy, there is more supply than demand. The consumer dictates the kind of products to be produced.

over the Internet. Fingar [1] described that technology and business move abreast of one another. Based on the changes in information technology, the timeline of business technology is divided into 6 stages. We are now entering the ubiquitous computing realm which focuses on Internet-based collaborative computing. The B2B interaction will be increasingly be digitised and migrated to cyberspace.

The new digital economy has transformed the business world from the traditional, linear value chain model to a centralised value network model. Figure 1 illustrates how information and product are passed within the traditional linear value chain model. The main drawback of this model is that the visibility of information is tier-dependent. Information is not properly shared between the parties upstream and downstream of the value chain. Even if information is shared, the information systems tend to be rigidly-configured, tightly-integrated, proprietary, non-reusable and unportable, often requiring high infrastructural set-up and maintenance costs. The chain, and hence the information systems to be used, tends to be dominated by the chain master.

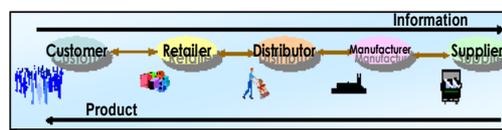


Figure 1 Traditional Linear Value Chain

Figure 2 illustrates a centralised value network where partners form a trusted network community to conduct collaborative commerce via a hub-and-spoke e-marketplace that facilitates the business processes. For example, Covinsit's [2] e-marketplace model transforms the linear chain into a far more efficient network model that facilitates information sharing within the community. However, the infrastructural cost is still substantially

high, and information systems remain tightly integrated and proprietary. Furthermore, the interconnectivity within the community is static in nature.

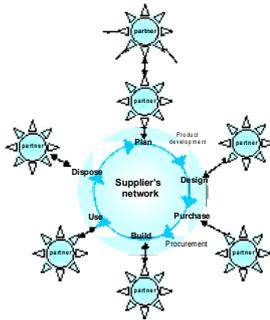


Figure 2 Centralised Value Network Model

In this model, the application service provider has to bear the full cost of the infrastructure that needs to be put in place to provide the service to the community. The cost of enterprise application integration (EAI) to the partners' back-end systems will be high as well, if they choose to have it. The interconnectivity within the community is static because it has to be pre-determined.

3. THE VIRTUAL COMMUNITY

Figure 3 shows an example of the complex, interwoven digital commerce connectivity among different clusters of value network communities (VNCs) that is beginning to emerge. This is a decentralised VNC model. The clusters of VNCs give rise to the dynamic nature of the digital economy. Some of the key characteristics of the digital economy are described as follows:

- each VNC consists of both supply and demand chains that form a value chain driving from customer demand to the delivery of the final product to the customer
- one VNC competes with another VNC
- a trading partner of a community can be a partner of another community
- each trading partner focuses on its core competencies

The success of a VNC requires the sharing of information, business processes and other resources among partners upstream and downstream of the individual value chains. It is therefore vital that partners in the community are virtually integrated with one another. Competing in the digital economy is neither just about technology nor just about business. It is about getting all the value chain partners to work together seamlessly, collaboratively, securely, dynamically as a digital commerce community over cyberspace, and hence, the formation of "Virtual Enterprises". In this context, the "virtual enterprise" comprises all trading partners, customers and suppliers working together as ONE single, extended entity.

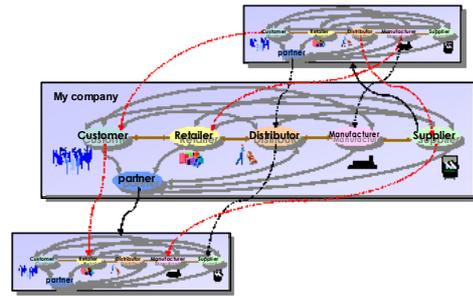


Figure 3 Decentralised VNC

4. CURRENT CHALLENGES

For enterprises, both large and small, to function as a member of a VNC, they are confronted with a number of business and technical challenges. The business challenges include:

- Ability to re-configure itself rapidly
- Lack of common standards in digital commerce exchange
- Ability to handle a diverse, de-centralised and dynamic collaborative environment
- Ability to provide visibility of timely business information

The technical challenges faced relate to making sense of the diversity of technologies and the multitude of applications being used by the members of the community, including

- Diverse IT applications for Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), Supply Chain Management (SCM), Product Design Management (PDM) and other customised applications.
- Diverse programming languages used for the implementation of IT applications.
- Diverse operating system and middleware platforms.
- Diverse networks types and topologies.

The key challenge is, therefore, to connect the islands of technologies and applications for information visibility and the formation of dynamic B2B relationships and connectivity for the network community.

5. VIRTUAL INTEGRATION INFRASTRUCTURE & FRAMEWORK

Reddy [4] defines virtual integration as being based on the enterprise infrastructure that provides the necessary mechanism for integrating a broad range of forward and backward activities within the value chain, including interaction with trading partners, industry networks, customers, and other maintenance and commodity suppliers. The virtual integration IT infrastructure would then consist of a collection of systems that create an Internet-based marketplace for buyers and sellers to transact in a VNC. Such an infrastructure should enable efficient sharing of information and resources as well as

seamless collaboration of trading partners with one another.

It is observed that three major factors need to be considered to achieve virtual integration for a VNC of trading partners, and they are:

- (1) A good IT framework that is able to support e-services, connectivity, interoperability, security, standardisation of business dictionary, inter-enterprise business process, scalability, reliability and flexibility.
- (2) A suitable business model for deploying the IT framework.
- (3) Commitment from top management.

Obviously, satisfying the first factor would put an organisation at an economic advantage as it would have a comprehensive IT framework to conduct on-line business with partners in the VNC. However, not all partners will have the luxury of owning such a framework because of the high cost of investment. There is, therefore, a need for a cost-effective IT infrastructure that can serve as the necessary framework for partners in a VNC to collaborate seamlessly over the Internet, which is proposed in the ensuing sections of this paper.

It is equally important for a company to have full commitment from the top management to ensure streamlining of the company's Internet strategy with its business strategy. Leading strategy experts Robert and Racine [4] proposed a comprehensive framework that connects a company's Internet strategy with the company's business strategy.

6. COMMON E-BUSINESS COMPUTING MODELS

The widely used computing models used to conduct e-business over the Internet are summarised as follows:

- *The centralised web-based architecture* model provides all on-line business applications for a trading partner community at a central location. A central web server is responsible for handling incoming user requests. Upon receiving a request, the web server gathers data from back-end applications, processes the data based on the appropriate business logic and responds to the user with the required information. In this model, the application can be custom-built or procured, and hosted within the trading partner company's premises. This is an attractive solution if the cost of IT investment is not a constraint of the company because it allows for total control and customisation of the application. However, it requires a substantial level of manpower resources to coordinate and execute implementation strategies, as well as maintain and update the systems.
- *Application Service Providers (ASPs)* offer an alternative solution to the centralised web-based model. ASPs essentially put their applications up for "rent" via the Internet to companies that need them. It is a cost-effective model for companies that have limited IT budget and resources. By paying for the use of such

remotely hosted applications, rather than procuring or developing, and maintaining them in-house, the total cost of ownership for a company could be substantially reduced. The ASP provides the majority of the IT infrastructure and manages the hardware and software upgrades. Such a model enables trading partner companies to "try out" newer technologies at lower costs, as there would not be a need for an upfront investment on hardware and software. However, the ASP would need to have a substantial client base to sustain its high set-up cost, operations cost and maintenance cost, and a typically high service availability rate.

The use of the Internet for e-Commerce has, to date, largely been based on the centralised, multi-tier computing model, such as the two described above. Typically, the functionality of such systems is broken up into three separate tiers, namely, the presentation, business logic and the data source tiers. Such a model may not fit well in the new era of digital commerce due to the complex business interfaces among trading partners. It also does not facilitate information to be shared easily among the VNC due to the complexity involved in B2B integration and the diverse technologies and information systems used by individual partner companies.

7. IMPACT OF PEER-TO-PEER ARCHITECTURE (P2P)

What is *peer-to-peer*? Briefly, it is an architecture model for that can be used to better utilise resources that are connected to a network, such as the Internet. It facilitates the sharing of data, computational resources and services by computers and devices residing on the network.

The main characteristics of P2P are outlined as follows:

- Peers understand both *requests* and *responses*.
- A peer can either act as a *client* or a *server*. It is a client when making a request and a server when responding to one.
- Peers are interconnected with one another via a network, such as the Internet.
- There is no single point of failure.
- Information and resources remain at the edge of the network.

The P2P architecture abstracts traditional, hierarchical networks into a set of nodes that are interlinked based on a set of well-defined rules that provide increased flexibility and mitigates traditional bottlenecks and points of failure. Presently, P2P networks are largely server-mediated. Each network contains both peers and a handful of central servers. These servers are responsible for maintaining a registry of the peers on the network so that they can locate one another. Multiple instances of these servers are usually present for the purpose of ensuring performance and providing redundancy in the event of failure of one or more servers.

In the dynamic digital economy, collaborative product commerce in a VNC is approaching that of a tier-less model. Transactions are being carried out between partner endpoints at the edge of the network community. Each partner endpoint can perform a dual role, one of being a commerce services *provider* as well as commerce services *consumer*. This relates well with the P2P computing model where clients can be servers and servers can also be clients.

From this close correlation between the characteristics of the VNC and those of the P2P computing architecture, it is apparent that the latter would potentially be a good fit for the former. Moreover, P2P overcomes the limitations of the centralised-server computing model by allowing trading partners to provide information services directly to one another. An example of the P2P computing model is Gnutella [7], an open, decentralised network that uses P2P technology for file sharing over the network. Figure 4 illustrates how Gnutella works.

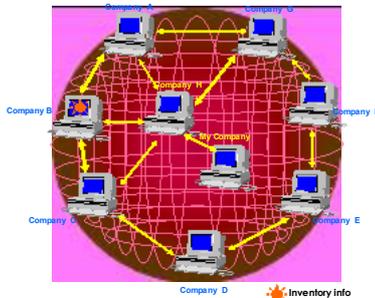


Figure 4 Gnutella Network

When My Company searches for inventory information located in Company B, the search is transmitted to everyone connected in the Gnutella network horizon. Although My Company is not directly connected to Company B, as long as the latter has allowed all partners in the community to view its inventory information, the former will be able to access it through one of the peers that is connected to the latter. Common characteristics of such a network are:

- Any peer can *publish* any information one wishes to share in the network.
- Any peer can *discover* new information in the network.
- Any peer can *consume* the required searched result.
- Any peer can be an information *provider* as well as an information *consumer*.
- New peers can easily *connect* to the network.

P2P technology is an ideal candidate to model the real-world dynamic connectivity required of the virtual integration among partners in a VNC. Each peer in the VNC represents a company that provides its core business functions as e-Services for its partners within the community to consume.

Mirroring the characteristics of the P2P network architecture, the key benefits of a P2P-enabled collaboration community are:

- Each partner in the community can be a service provider as well as a consumer.
- The ability to publish new information to the community as and when it is available.
- The ability to discover and consume new information introduced in the community.
- The ability to rapidly integrate the trading partners in the community.

However, there is a drawback imposed by the existing P2P architecture in that although new information being published to the network can be automatically discovered and consumed, the same cannot be said of services. When a new service has been published to the community, each peer would have to be notified and reconfigured for it to be able to consume the new service. There is no standardised way of describing services and how a peer can be configured to consume new services. One way to resolve this is to leverage Web Services technology to enhance the P2P architecture.

8. WEB SERVICES

Web Services are software components that are programmatically accessible as services over the Internet. The platform- and language-independent interfaces of Web Services allow for easy integration of diverse applications residing on heterogeneous systems. Web Services focus on moving data around using XML [6] to keep inter-application communication loosely-coupled. The publication, discovery and invocation of Web Services over the Internet make use of the following open standards:

- **Web Service Description Language (WSDL)** [8] – a standard that describes a Web Service type and its interfaces.
- **Universal Description, Discovery and Integration (UDDI)** [9] – a standard mechanism for the registration and discovery of Web Services.
- **SOAP (formerly an acronym for Simple Object Access Protocol)** [7] – a standard messaging protocol for Web Services.
- **Extensible Markup Language (XML)** [6] – defines the structured data to be interchanged among applications.

Web Services and P2P complement each other to provide an infrastructure to support the virtual integration of partners within a VNC to enhance their collaborative commerce activities. Figure 5 depicts the conceptual roles and operations of peers in a Web Services-based environment. It illustrates the way Web Services and P2P work side-by-side to complement each other.

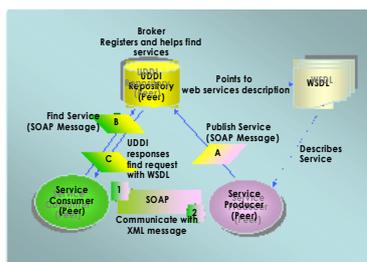


Figure 5 Marrying Web Services & P2P

There are three basic roles. They are the service *provider*, the service *consumer*, and the service *broker*. The service provider makes available the service contract and registers (A) the service with a service registry. A service consumer queries the registry, which serves as the broker, and finds (B) a compatible service. The service registry gives the service consumer directions on where to the service is located and its relevant contract (C). The service consumer uses the contract obtained to bind ((1) and (2)) the client to the service.

The service provider can be a service consumer, and vice versa (for example, in the case of service aggregators). The service registry's responsibility is to maintain a central repository or directory of information on services that have been published for consumption.

9. BENEFITS OF COMBINING WEB SERVICES AND P2P

The use of Web Services and P2P technologies for virtual integration will give rise to the following potential benefits:

- As it is a distributed architecture, the cost is shared among the partners of the VNC, each bearing a fraction of the infrastructure cost. This is in contrast to the centralised web-based architecture where a single party bears the cost and controls the applications that serves the community.
- VNC members can concentrate on providing core business functions as web services for partners to discover and consume, as and when needed.
- VNC members can dynamically discover, establish and conclude short-term trading partnerships without substantial integration overheads.
- VNC members have the flexibility to add new services or update existing services as and when needed, at the edge of the network without affecting other peers.
- The VNC can share resources, for example, by not duplicating existing services that are already being offered by a member.
- VNC members own their business data and maintain their own applications. There is no concern of data privacy as opposed to that of a centralised approach.
- By allowing the community as a whole to better utilise and manage its resources, it is able to achieve better efficiencies to compete with rival communities.
- Information can be pushed as well as pulled, on-demand when required.

- Scalability is enhanced for the VNC as a whole, and single points of failure and other performance-related issues are reduced.

10. CONCLUSION

The combination of Web Services and P2P provides an ideal platform for the virtual integration of partners in a VNC. Using such a platform, individual trading partners have the freedom to make and break short-term partnerships without having to incur large overheads required in integrating with each partner they collaborate with. Data and services provided within the community are always up-to-date as they are put out by the members themselves instead of a central service provider. Data privacy concerns are also eliminated with each member of the community hosting its own data and services instead of making use of a centralised service provider. Performance issues are mitigated by the distributed nature of the services, resulting in a more efficient VNC as a whole.

11. FUTURE RESEARCH

Despite all the benefits described earlier, there remain a number of gaps to be filled before the full benefits of virtual integration can be realised. They include:

- The definition of an inter-enterprise public business processes standard. This to ensure a common set of business dialogues for business process integration between partners in a community.
- The definition of standard business and technical dictionaries to be used for terminologies to be exchanged in documents and messages in business transactions over the network. This is to ensure that a common business language is used between trading partners for on-line B2B transactions.
- The definition of an architectural layer for business process management (or BPM) that enables the orchestration of business processes across the various stages of the life cycle of a project.

REFERENCES

- [1] Fingar, Peter & Aronica, Ranald, *The Death of "e" and the Birth of the Real New Economy*, Meghan-Kiffer Press, 2001.
- [2] Raisch, Warren D., *The e-Marketplace Strategies for Success in B2B eCommerce*, McGraw Hill, 2001.
- [3] Reddy, Ram & Reddy, Sabine, *Supply Chain to Virtual Integration*, McGraw Hill, 2001.
- [4] Robert, Michel & Racine, Bernard, *e-Strategy Pure & Simple*, McGraw Hill, 2001.
- [5] <http://www.gnutella.com/>
- [6] <http://www.w3c.org/XML/>
- [7] <http://www.w3c.org/2000/xp/Group/>
- [8] <http://www.w3c.org/2002/ws/desc>
- [9] <http://www.uddi.org/>
- [10] <http://www.rosettanet.org/>

A Study of Behavioral Intention for Mobile Commerce Using Technology Acceptance Model

Yuchung Hung¹, Haoerl Yang², Chinghung Hsiao¹, Yunglu Yang¹

¹Department of Information Management, Tatung University, Taipei, Taiwan, China

²Department of Business Management, Tatung University, Taipei, Taiwan, China
{¹hung6599;²heyang}@ttu.edu.tw

ABSTRACT

Davis's Technology Acceptance Model (TAM) suggests that perceived ease of use and perceived usefulness can predict user behavior and acceptance when using information technology. In this study we use it as a tool for evaluating the acceptance and usage for mobile commerce by employing data collected from 205 students selected proportionally from nine universities located at Taipei City. The results show that perceived ease of use and perceived usefulness are the fundamental determinants of user acceptance and have positive impact. They demonstrate positive impacts on attitude to use and behavioral intention to use. In addition, the attitude toward using m-commerce has a partial mediating effect on behavioral intention to use. Behavioral intention to use exerts a positive, though lesser, impact on actual use of m-commerce.

Keywords: Technology Acceptance Model, M-commerce, Confirmatory Factor Analysis, Regression Analysis, partial mediating effect.

1. INTRODUCTION

Digital mobile phone has improved the defects of analogue mobile phone, and brought mobile communication into a digital era. Governments of many countries have released their telecom policies for new dealers to invest in the market so as to compete with each other and cause the telecom fee to reduce. In the meantime, emulation among the dealers will bring innovations of communicative semi-conductor technique. As a result, the "used to be" heavy mobile phone now turns into portable and light personal stuff, which makes the mobile communication service grow rapidly.

Ever since the "i-Mode" service promoted by NTT DoCoMo, all global wireless dealers pay great attention to this emerging commerce, and start to run a similar business model which leads to a trend for crashing in the wireless network. With the vast application of Wireless Application Protocol (WAP), the promotion of GPRS communication service, and various newly promoted wireless terminal equipments (such as Smart Phone, Wireless PDA etc.), the integration of wireless communication and network will quicken the grand opportunities.

The technology acceptance model (TAM) proposed by Davis (1989) has been widely applied and empirically tested in the adoption or acceptance of information technology (IT) to explain end-users' acceptance

behavior. The technology such as e-mail and graphics, word process, voice mail were tested with TAM to explain the users' acceptance behavior. However, a meta-analysis of empirical findings indicates that the relationship between ease of use and acceptance is weak and the relationships in the model may be moderated by

a third variable such as gender, culture (Ma and Liu, 2004). In this study, we use m-commerce technology to test if the relationships in the TAM still validate.

2. LITERATURE REVIEW

2.1 Technology Acceptance Model

Davis (1986) used belief-attitude-intention-behavior relationship of TRA (Theory of Reasoned Action) to develop TAM to model user acceptance of IT, as shown in Figure 1. This model provides a theoretical foundation to understand how external variables influence the user's inner beliefs in a technology, and then predict his/her behavior, which in turn affect the use of technology. The purpose of TAM is to provide an explanation toward the acceptance of technology which explains user's behavior on accepting new IT, and analyze the factors that influence their attitude toward using new IT. The design of the constructs used in TAM is both valid and reliable, and has been discussed, experimented, and amended by many scholars. So the same constructs could be adopted even though the technologies to be studied are different. (Adams et al., 1992; Taylor and Todd, 1995; Mathieson and Chin, 2001).

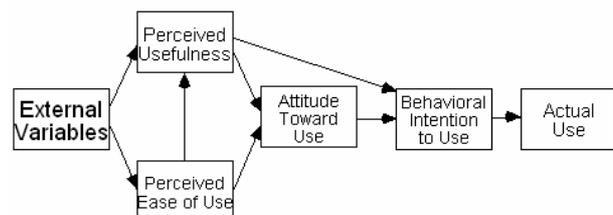


Figure 1 Technology acceptance model
(Davis et al., 1989)

2.2 Mobile Commerce

Durlacher (1999) figured mobile commerce (m-commerce) as a vast area of activity comprised of transactions with monetary value conducted via wireless network. Other scholars like Skiba et al. (2000) defined m-commerce as transactions using a wireless device either in word or speech that result in the transfer of value in exchange for information, services, or goods. Jiang (2001) mentioned that m-commerce is the transactions conducted by terminal equipment, such as PDA, and SmartPhone etc, and the services should be paid. The related services and application include the downloading of music and picture, mobile monetary control, and mobile advertising.

From the above definitions, we can find that the divarication for the definition of m-commerce is whether monetary transactions should be included. The definition by Skiba et al. (2000) will be adopted as the foundation for this study, which referred m-commerce to transactions using a wireless device either in word or speech that result in the transfer of value in exchange for information, services, or goods.

The definition for m-commerce by Skiba et al. (2000) is not restrained for monetary service only. Therefore, the formation of m-commerce will depend upon the match of software and hardware. According to the framework proposed by Varshney and Vetter (2002), the formation of m-commerce is divided into two main sections, application, and development and supply. For development and supply, there are three categories: the supplier of service, the supplier that provides content, and the developer of application framework. Therefore, there are four levels for application, infrastructure for wireless network, intermediary software, device, and the application of m-commerce. This framework provides the m-commerce not only a direction but also a reference for standard allocation before investing in m-commerce.

3. RESEARCH MODEL AND HYPOTHESES

This study adopts the extended structural definition of TAM proposed by Davis et al. (1989), without considering external variables, as shown in Figure 2.

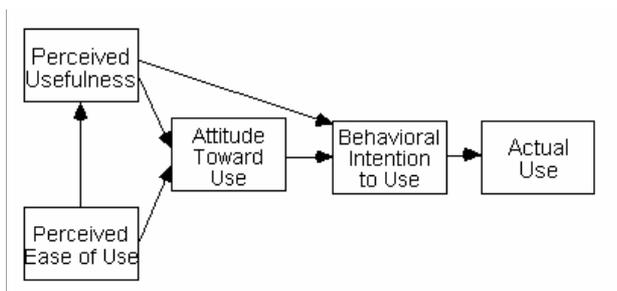


Figure 2 Research framework

3.1 Measures

In this study we used five questionnaire items to describe the construct of perceived usefulness and four items for perceived ease of use, which were adapted from Davis (1989) and Moore and Benbasat (1991) and measured on a five-point Likert scale with anchors ranging from strongly agree (1) to strongly disagree(5).

Attitude toward using m-commerce were indicated by four items adapted from Ajzen and Fishbein (1980), Malhotra and Galletta (1999) and measured on a five-point Likert scale; behavioral intention to use by two items adapted from Ajzen and Fishbein (1980), Davis (1989) on a five-point Likert scale.

As for the actual use of m-commerce it was defined as the frequency used per week (Malhotra and Galletta, 1999; Igarria et al., 1997).

3.2 Research Hypotheses

The research hypotheses for this research were made according to the structural theory of TAM proposed by Davis et al. (1989).

H1: Perceived ease of use will have a positive effect on perceived usefulness.

H2: Perceived usefulness will have a positive effect on attitude toward using m-commerce.

H3: Perceived ease of use will have a positive effect on attitude toward using m-commerce.

H4: Perceived usefulness will have a positive effect on behavioral intention toward using m-commerce.

H5: The attitude toward using m-commerce will have a positive effect on behavioral intention of use.

H6: The behavioral intention to use will have a positive effect on actual use of m-commerce

3.3 Data Collection

According to the ten years information communication technology research for university students made by Macleod et al. (2002), the percentages of students that were expected to use information communication technology frequently increased from 15%~30% to 65%, and the percentages of emphasize on self-use also increased from 20%~35% to 73%, which showed that most students had more desire to use information communication technology. According to the 2004 Taiwan Internet and Broadbar Usage Survey published by TWNIC, as of the end of January 2004, they are 12.64 millions Internet users in Taiwan and the number of broadband Internet users are 9.03 millions and about 81.55% with age of 21~25 (TWNIC, 2004). Since almost of those users are college students, they can be formed the sampling units. However, due to the limited resources, we decided to collected data from the students studied at nine universities located in Taipei City. We planned to select a sample from each university with

sample size proportional to its size. So 500 questionnaires were delivered with 205 usable copies returned including 95 female and 110 male.

3.4 Analytical Procedure

In this study we used Cronbach's α with the cut-off of 0.7 (Nunnally, 1994) to assess the reliability of each construct. Exploratory factor analysis using principle component analysis with varimax rotation and confirmatory factor analysis were employed to assess construct validity and unidimensionality. The former is to investigate if the number of factors extracted is the exact number of constructs determined prior to the analysis and the later to assess construct validity and unidimensionality.

The Correlation analysis and multiple regression analysis were used to test the relationships among the scales or constructs and the effects on dependent variables. The causal direct, indirect, and total effects between the exogenous and endogenous variables can be calculated by using structural equation modeling (SEM).

4. DATA ANALYSIS

4.1 Reliability and Validity Analysis

The Cronbach's α coefficients for perceived usefulness, perceived ease of use, attitude toward using m-commerce, behavioral intention to use m-commerce, are 0.8969, 0.8028, 0.9216, and 0.7562 respectively. All those Cronbach's α 's are above the cut-off of 0.70. The reliability of each construct is acceptable.

Exploratory factor analysis for the items of perceived usefulness and perceived ease of use reveals that the items produce two proposed factors with total variance explained 68.7% and the individual items are loaded with loadings above 0.7 on their appropriate factors as intended. The items for attitude toward using m-commerce, behavioral intention to use reveals and actual use produce three proposed factors with total variance explained 83.9% and all loadings above 0.78. The results of confirmatory factor analysis also show the consistency with the findings in exploratory factor analysis. The validity and unidimensionality for the constructs of both perceived ease of use and perceived usefulness are supported. The validity and unidimensionality for the constructs for attitude toward use, intention of use, and actual use are also supported.

4.2 Correlation Analysis

The averaged item scores are used as a measure of their posited construct. The correlation matrix for the constructs used in the model are shown in Table 1. It indicates that the correlation between each pair of constructs is statistically significant at the 0.05 level, except the pair of behavioral intention to use (*intention*)

m-commerce and actual use (*action*) of m-commerce (marginal significance, p-value=0.061). Independent variables, perceived usefulness (*use*) and perceived ease of use (*ease*) show significantly positive correlation with attitude toward use (*attitude*), 0.656 (p=0.000) and 0.543 (p=0.000) respectively, and with behavioral intention toward use, 0.527 (p=0.000) and 0.457 (p=0.000) respectively, at $\alpha=0.01$. The correlation coefficient between perceived usefulness and perceived ease of use (0.477) is significantly positive at $\alpha=0.01$. The correlation coefficient between attitude toward use and behavioral intention to use (0.591) is significantly positive at $\alpha=0.01$.

Table 1 The correlation analysis of variables

	<i>use</i>	<i>ease</i>	<i>attitude</i>	<i>intention</i>	<i>action</i>
<i>use</i> Sig.	1 .				
<i>ease</i> Sig.	.477** .000	1 .			
<i>attitude</i> Sig.	.656** .000	.543** .000	1 .		
<i>intention</i> Sig.	.527** .000	.457** .000	.591** .000	1 .	
<i>action</i> Sig.	.224** .001	.201** .004	.162* .021	.131 .061	1 .

** : Correlation is significant at the 0.01 level (2-tailed)

* : Correlation is significant at the 0.05 level (2-tailed)

4.3 Hypotheses Testing

In correlation analysis the correlation coefficient between perceived ease of use and behavioral intention to use (0.457) is significant with p-value=0.000. Regress behavioral intention toward use on perceived usefulness, perceived ease of use and the attitude toward using m-commerce with stepwise method. The results, as presented in Table 2, show that the three independent variables are all significant in predicting the dependent variable with standard beta coefficients in the regression equation, $b_{Attitude}=0.368$, $b_{Use}=0.211$, $b_{Ease}=0.157$. The total variance of intention toward use can be explained 0.391 (adjusted R^2) by perceived usefulness, perceived ease of use and attitude toward use

Table 2 Stepwise regression analysis

Order	R	A- R^2	Δ A- R^2	Std. β	Δ F	Sig.
1. <i>attitude</i>	0.591	0.346	0.346	.368*	109.058	0.000
2. <i>use</i>	0.619	0.377	0.031	.211*	11.141	0.001
3. <i>ease</i>	0.633	0.391	0.014	.157*	5.581	0.019

Dependent variable: *intention* (Behavioral intention to use)

*significant at $\alpha=0.05$, A- R^2 : Adjusted R^2 .

In order to test the hypotheses we established the related regression models as presented in Table 3. All models are significant at $\alpha=0.05$, except the model *action=intention*, significant at $\alpha=0.10$.

Table 3 Regression analysis

Model	Variable	β	Std. β	A- R ²	Sig.
use=ease	Constant	2.131		0.224	0.000**
	<i>ease</i>	0.448	0.477		
attitude=Use+Ease	Constant	0.479		0.494	0.000**
	<i>use</i>	0.597	0.514		
	<i>ease</i>	0.324	0.297		0.000**
	Constant	0.985		0.377	0.001**
intention=use+attitude	<i>use</i>	0.295	0.244		
	<i>attitude</i>	0.448	0.431		0.000**
action=intention	Constant	1.673		0.012	0.061*
	<i>intention</i>	0.153	0.131		

Table 4 shows that the results of the six hypotheses testing in this research, which indicates that all research hypotheses are supported by data.

Table 4 The results of research hypotheses testing

Item	Hypotheses	Result
H1	Perceived ease of use positively influences perceived usefulness.	supported
H2	Perceived usefulness positively influences attitude toward using m-commerce.	supported
H3	Perceived ease of use positively influences attitude toward using m-commerce.	supported
H4	Perceived usefulness positively influence behavioral intention toward using m-commerce.	supported
H5	The attitude toward use positively influences on behavioral intention toward using m-commerce.	supported
H6	The intention toward use positively influences actual use of m-commerce.	supported

The causal direct, indirect, and total effects between the exogenous and endogenous variables were calculated by using SEM. As can be seen in Table 5, all the effects are statistically significant at or above the 90% confidence level. For example, the total and indirect effects of perceived ease of use on attitude toward use are .543 (t value=8.48) and .246(t value=5.72), respectively. The direct effects presented in Table 5b can be shown in TAM diagram as in figure 3.

Table 5 Total effects, direct effects, and indirect effects

a. Standardized total effects (t value)

	<i>ease</i>	<i>use</i>	<i>attitude</i>	<i>intention</i>
<i>use</i>	.477 (6.91)	.000	.000	.000
<i>attitude</i>	.543 (8.48)	.514 (6.76)	.000	.000
<i>intention</i>	.350 (5.47)	.466 (7.17)	.431 (5.32)	.000
<i>action</i>	.046 (1.70)	.061 (1.79)	.057 (1.78)	.131 (1.93)

b. Standardized direct effects (t value)

	<i>ease</i>	<i>use</i>	<i>attitude</i>	<i>intention</i>
<i>use</i>	.477 (6.91)	.000	.000	.000
<i>attitude</i>	.297 (4.24)	.514 (6.76)	.000	.000
<i>intention</i>	.000	.244 (2.77)	.431 (5.32)	.000
<i>action</i>	.000	.000	.000	.131

	<i>ease</i>	<i>use</i>	<i>attitude</i>	<i>intention</i>
				(1.93)

c. Standardized indirect effects (t value)

	<i>ease</i>	<i>use</i>	<i>attitude</i>	<i>intention</i>
<i>use</i>	.000	.000	.000	.000
<i>attitude</i>	.246 (5.72)	.000	.000	.000
<i>intention</i>	.350 (5.47)	.222 (4.35)	.000	.000
<i>action</i>	.046 (1.70)	.061 (1.79)	.057 (1.78)	.000

The t values >1.66, 1.98, and 2.35 are considered to be significant at the 0.10, 0.05, and 0.01 levels, respectively.

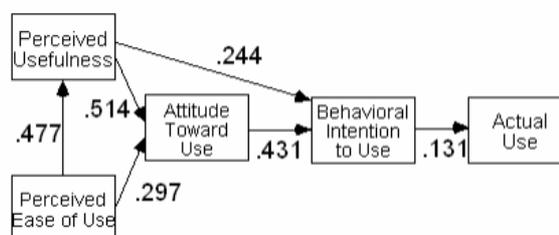


Figure 3 The causal direct effects between constructs

5. CONCLUSIONS AND SUGGESTIONS

A number of hypothesized relationships in Davis's TAM with different IT, i.e. m-commerce, were tested based on a sample of 205 college students. The results of our study generally confirm the Davis's findings. The relationships between perceived ease to use and perceived usefulness, and between perceived usefulness and attitude toward use are strong. The relationships between perceived ease to use and attitude toward use is also strong, while weak reported by Ma and Liu(2004). We found from regression analysis that behavioral intention to use can be predicted by perceived usefulness and attitude toward use.

Although we did not proposed a formal hypothesis about the mediating effect of attitude toward use, we added one more regression model,

$$intention = .400 use + .266 ease,$$

p values=.000 for beta testing,

with two models, that is, $attitude = .514 use + .297 ease$, p values=.000, from Table 3, and $intention = .211 use + .157 ease + .368 attitude$, p values< .050, from Table 2. By mediating effect test (Baron and Kenny, 1986) the beta coefficient of *use* decreases from .400 to .211 when adding *attitude* toward use into the regression model, which demonstrates that attitude toward use has a partial mediating effect on behavioral intention to use. However, the effect of behavioral intention to use on actual use is marginally significant (p=0.061).

In this survey one question about the most common used

m-commerce service, it is the message-sending, the most typical service provided. About the service the users of m-commerce expect most is the improvement of entertainment service, followed by global positioning, financial service, and general information. The above information implies the trend that the users may gradually accept more m-commerce services and use them.

That the effect of behavioral intention to use on actual use is statistically marginal significance may imply that there are some important variables omitted, such as the service charge, the security of the system, and the problem of on-line quality, when users considering taking action to use.

Because m-commerce is expected to have a prosperous market for service providers, they should realize user's behavior in the market and manage potential user's requirements for it. Potential user's attitude toward using m-commerce and the factors that influence their intention to use should be realized and emphasized in order to develop proper marketing strategies for the use of m-commerce.

REFERENCES

- [1] Adams, D.A., R.R.Nelson, and P.A. Todd, , "Perceived Usefulness, Ease of Use, and Usage of Information Technology: A Replication", *MIS Quarterly*, Vol.16, No.2, pp277-247, June 1992.
- [2] Ajzen, I. and M. Fishbein, *Understanding Attitudes and Predicting Social Behavior*, Prentice-Hall, Englewood Cliffs, 1980.
- [3] Baron, R.M. and D.A. Kenny, "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations", *Journal of Personality and Social Psychology*, Vol. 51, No. 6, pp 1173-1182, 1986.
- [4] Davis, F. D., "A Technology Acceptance Model of Empirically Testing New End-User Information Systems: Theory and Results", *Doctoral Dissertation, Sloan School of Management, Massachusetts Institute of Technology*, 1986.
- [5] Davis, F. D., "Perceived Usefulness, Perceived Ease of Use and User Acceptance of Information Technology", *MIS Quarterly*, pp319-340, Sep. 1989.
- [6] Davis, F.D., R.P. Bagozzi and P. R. Warshaw., "User acceptance of computer technology: A comparison of two theoretical models", *Management Science*, Vol.35, No.8, pp982-1003, August 1989.
- [7] Durlacher, "M-commerce Report ", *Durlacher Research Ltd*, <http://www.durlacher.com>, 1999.
- [8] Igbaria, M., N. Zinatelli, P. Cragg, and A.L.M. Cavaye, "Personal Computing Acceptance Factors in Small Firms: A Structural Equation Model", *MIS Quarterly*, Vol.21, No. 3, pp 279-305, 1997.
- [9] Jiang, Pei-Wei, "The growth estimation of global m-commerce market", *MIC Market Intelligence Center*, <http://mic.iii.org.tw/intelligence/>, May 2001.
- [10]Ma, Qingxiang and Liping Liu, "The technology Acceptance Model: A meta-Analysis of Empirical Findings" , *Journal of Organizational and End User Computing*, 16, 1, Jan-Mar 2004.
- [11]Macleod, H., D. Haywood, J. Haywood, and C. Anderson., "Gender and information and communications technology_ a ten year study of new undergraduates", *TechTrends*, Vol. 46, pp 11-15, 2002.
- [12]Malhotra Y. and D. F. Galletta, "Extending the Technology Acceptance Model to Account for Social Influence: Theoretical Bases and Empirical Validation", *Proceedings of the 32nd Hawaii International Conference on System Sciences*, 1999.
- [13]Methieson K., E. Peacock, and W. Chin, , "Extending the Technology Acceptance Model: The Influence of Perceived User Resources", *DATA BASE for Advances in Information Systems*, Vol. 32, No.3, pp86-112, 2001.
- [14]Moore, G. C. and I. Benbasat, "Development of an instrument to measure the perception of adopting an information technology innovation", *Information Systems Research*, Vol.2, No.3, pp192-222, 1991.
- [15]Nunnally, J.C., *Psychometric Theory*, 3rd ed., New York: McGraw-Hill, 1994.
- [16]Skiba, Brian, Johnson, Mairi, Dillon, Michael, "Moving in Mobile Media Mode", *Lehamn rothers*, <http://www.entsoftware.com>, 2000.
- [17]Taylor, S. and P. Todd, , "Assessing IT Usage: The Role of Prior Experience", *MIS Quarterly*, Vol.19, No.4, pp561-570, December 1995.
- [18] TWNIC(*Taiwan Network Information Center*),.. http://www.twnic.net.tw/file/twnic_summer.pdf 2004.
- [19]Varshney, U. and R. Vetter, "M-commerce: Framework, Applications and Networking Support", *Kluwer Academic Publishers*, Vol.7, pp185-189, 2002.

Ambient E-Services: Framework and Applications

Yuanchu Hwang^{1,2}, Soetsyr Yuan¹

¹ Department of Management Information Systems, National Chengchi University, Taiwan, China

² Department of Information Management, Dahan Institute of Technology, Taiwan, China
{ychwang, yuans}@mis.nccu.edu.tw

ABSTRACT

Most of existing mobile services were designed based on the client/server architecture. Those mobile services neither paid much attention to mobile users' interactions with their environments nor considered the cooperation efforts between the mobile users in a dynamic peer group. In this paper, the notion of ambient e-services is defined to identify a new scope of mobile e-services. The notable features of ambient e-services are the exhilarated linkage based on social context and significantly rapid growth of connections. We also present an ambient e-service framework that characterizes ambient e-services with three dimensions (value stack, environment stack and technology stack). We exemplify several ambient e-service applications, which rest on the mobile peer-to-peer technology and ambient context aware sensors environments. Ambient e-services make a mobile user not only connect to dynamic ambient environments but also build channels that connect with other mobile users in the nearby surrounding environment, capitalizing dynamic environmental values as well as dynamic social values. Lastly, we identify for ambient e-services certain important issues worthy of future intensive research.

Keywords: Ambient e-Services, Mobile commerce, Context awareness, Social context, Mobile Peer-to-Peer

1. INTRODUCTION

Mobile commerce promises to deliver the real potential of Internet for commercial purposes to a significantly expanded market of existing and new users. A four-level integrated mobile commerce framework was proposed in [1], which discussed how to successfully define, architect, and implement the necessary hardware/software infrastructures in support of mobile commerce. However, the applications mentioned in [1] were grounded in the client/server architecture where the only interactions involved are between a services provider and a mobile user. Mobile users were standalone due to the lack of technology functions and communication channels that enabled mobile users to interact with each other; interactions or cooperation between mobile users were not considered as an important issue in mobile service scenarios. Consequently, the collected value attained from the mobile users of a peer group (or multiple peer groups) cannot be produced.

Fortunately, the latest Peer-to-Peer (P2P) technology (e.g., JXTA) makes it possible for individual mobile peers communicate with each other and wirelessly exchange their own information under sensors-enabled environments. One may notice that P2P computing has increasingly emerged at the forefront of Internet computing. The cooperative computing and resource sharing has been around for quite a while [2]. The collaborative interactions between mobile users bring mobile commerce into a new paradigm. Accordingly, a new framework of mobile applications should be investigated, considering dynamic environments and collected efforts of mobile users.

The ambient intelligence refers to the vision that technologies become invisible, embedded in our natural surroundings, present whenever we need them, enabled by simple and effortless interactions attuned to all our senses,

adaptive to users and context and autonomous acting [8].

In other words, ambient intelligence is to embed information representation in everyday objects and the physical environment becomes an interface to digital information. In this paper, we embrace this vision into mobile e-services built with the Mobile-P2P technology, unfolding the notion of ambient e-services.

Moreover, Location-based services (LBS) employ knowledge of the user's location to enable the provision of new or enhanced services to a user via a wireless handheld device (e.g. mobile phones, PDAs). Wireless e-services can be potentially very personal, timely and relevant, or even integrated with other services in a near-seamless way [4]. Within ambient environments, wireless handheld devices are very personal to a user and carried by the person; the context of the user (e.g., time and place) can be measured and interpreted; services can be provided at the point of need; and applications can be highly interactive, portable and engaging.

However, some people may wonder if current sensor networks can replace ambient e-services. Wireless Integrated Network Sensors (WINS) provide distributed network and Internet access to sensors, controls, and processors that are deeply embedded in equipment, facilities, and the environment. However, in sensor network environment, peers within the range may not know who others peers are (i.e., social context is confidential and presumed to be retained privately). But ambient e-services are capable of leveraging peers' private social context or past interactive experience, diversifying the potential opportunities of the dynamic collected efforts. There exists something about ambient environments that the sensors network cannot accomplish. The social connections based on the context aware environment provide a trustworthy linkage on the strength of the social networks and keep sensitive data from others. The social

contexts accordingly must be retained in the personal devices. Otherwise, sensitive data kept in a central server will makes user loath to participant in such services. Compare with the client/server architecture, the P2P design makes the connection numbers in a significantly rapid progression growth especially in an open space. Those are two distinguished characteristics that make the P2P design more appropriate than client/service architecture in ambient e-services.

This paper does not intend to offer fixed answers about how ambient e-service application systems should be designed in the future. Instead, it aims to motivate a new direction of research, which addresses challenges and research questions that arise when we take a system-oriented but high-level view of ambient e-services. We are interested in examining those novel applications and technologies that constitute the frontier of the ambient e-services.

The ambient e-service is a new thinking. To ward off ambient e-services framework limited to those scenarios we provided, we deploy the presentation of the paper in a top down structure. In section 2, we propose the ambient e-service framework that represents the ambient applications in three dimensions. Section 3 illustrates several key ambient e-service scenarios and their basic interaction architectures. Then we discuss those issues arise in ambient e-services in section 4. Finally, in section 5 we describe some conclusion remarks and future research topics.

2. AMBIENT E-SERVICES FRAMEWORK

In this section, we present an ambient e-service framework (as shown as Figure 1), identifying some possible deliverables (values) of ambient e-services and addressing the technologies required to support the applications of ambient e-services. The framework is composed of three dimensions, the value stack, the environment stack and the technology stack. The descriptions of the stacks will be detailed in the following subsections.

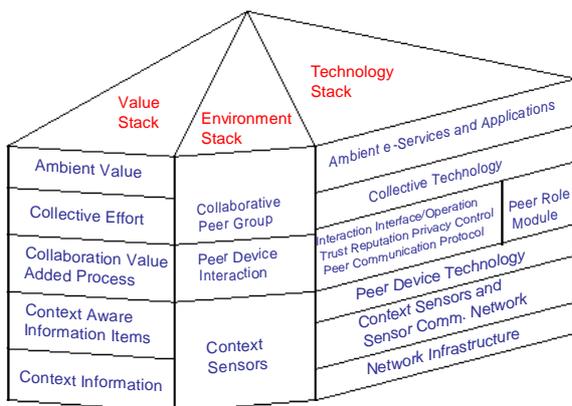


Figure 1. Ambient e-Service framework

2.1 Value Stack

The ambient value stack comprises five layers indicating the supporting value layers for ambient e-services

(deliverables of higher levels requiring the provision of deliverables of lower levels). The basic layer is the “Context Information”, which is attained from the ambient sensor environments. A mobile user can interact with the environments (e.g., entering a room) and the context sensors then communicate with the device the user carries (e.g., informing the user of his/her location). Comparing the user’s profile, the second layer ”Context aware Information Items”send to the user information items that matches his/her preferences.

The information items received from the environments may be useful to the user, but sometimes some of the information items may be irrelevant. With the communication ability, a mobile peer can collaborate with another mobile peer and exchange carried information items stored in their mobile devices. Through these collaboration interactions, it may be the case that there is someone whose information items are valuable to me and thus I can exchange or trade with them with a micro payment or a barter process. This is called “Collaboration Value Added Process”.

Not only communicating with only one peer, a peer can also interact with a peer group of multiple peers. Aiming at the same goal or interests, a peer group can collect their abilities to accomplish the goal that is impossible for a single peer to reach. Therefore, “Collective Effort” represents the power of a peer group (or multiple peer groups). The ambient value subsequently refers to the collective efforts, which include the peers’ interactions and peer groups’ interactions.

2.2 Environment Stack

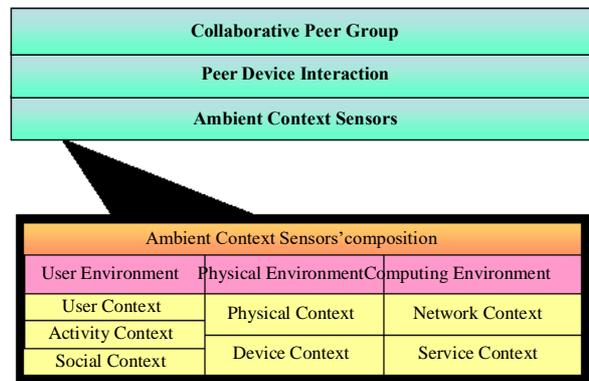


Figure 2. Ambient environment stack

The environment stack consists of three layers indicating the supporting environment layers for ambient e-services (environments of higher levels comprising those of the lower levels). The sketched stack diagram is shown in Figure 2.

The “Ambient Context Sensors”is the bottom layer, which include three categories of environments. Schilit, Adams & Want claim that the important aspects of context are: where the user is, who the user is with, and what resources are nearby. [12]

The user environment includes the user context, the activity context and the social context. The user environment represents the profile of a user, such as peer identification, where the user is, who the user is, the user's preference, their privacy concern, their social situation and relations with others, etc. The physical environment includes the physical context and the device context. The physical environment refers to things like temperature, noise and lighting level and device context. The computing environment includes the network context and the services context. The computing environment represents the network connectivity, available processors, cost of computing, bandwidth and available nearby services, etc.

The layer "Peer Device Interaction" represents environments (featuring peer-to-peer interactions) in which mobile users may exchange their experience and trade attained information items with nearby peers. The "Peer Device Interaction" layer rests on the "Ambient Context Sensors" layer; that is, all peers' information items originally come from the interaction with the dynamic environments of the bottom layer.

The top layer of the environment stack represents environments featuring peer groups' interactions and collaboration. Peer group can collect all peers' power for certain objectives (e.g., collective buy, forming a task-oriented workforce group, etc.).

2.3 Technology Stack

Network infrastructure is the fundamental of all ambient technology stacks. The wired and wireless hybrid network infrastructure provides the basic communication infrastructure for context sensors and peer devices. Above the network infrastructure layer, context sensors and sensor communication can interact with the surrounding peers. Peer devices function on top of the sensor communication network and the mobile device's hardware, operation systems and software platforms.

The layer above the peer device technology layer has two folds: (1) For peers to interact with each other, a peer communication protocol is required (similar to that in web services). (2) In ambient environments peers may just know nearby peers for quite a short period of time. How can a peer trust the nearby peers? Is it possible to build a reputation system for the aforementioned situation so that it is feasible to trust unfamiliar peers? Similarly, privacy control should be addressed in ambient e-services. Resting on trust, reputation and privacy control as the communication basics, mobile users can engage ambient interactions to handle some operations between peers. These operations are the likes of "discovery", "request", "respond", "bargain", "transact" and "delivery".

Since a mobile user plays different roles in different interactions, the peer role module can be applied to the peer in response to the dynamic roles rendered. Mobile users, who have the information items required by other users, play the seller role by applying the seller module. The surrounding peers (who have attained various kinds of information) then can play as the broker role by applying the broker module. Mobile users, who need the information

from others, apply the buyer module and play the buyer role (who can trade or barter with other mobile users).

Beyond the peer to peer interactions, peer group collective effort is the communal power of peers. Collaborative technologies that support peer and peer groups to work together underlie the key character of ambient power. Based on the collective technology, ambient e-services can be unfolded in various applications, which we will discuss in the next section.

3. AMBIENT E-SERVICE APPLICATIONS

It will be useful, to begin with, making a distinction between two kinds of ambient e-Services. One is for the distributed trading; another is for the distributed collaboration. Please note that the social context and the rapid growth of connections are the major incentives of applying the ambient e-service design. The two characteristics are not mutually exclusive; of course, one scenario may apply both of the ambient characteristics.

This section aims to identify several important ambient e-service application scenarios presuming ambient environments of mobile commerce are provided. Without loss of generality, the physical context of location is referenced and considered in the ambient environments addressed in the following scenarios, naturally relating to the LBS research. As mentioned in Section 1, ambient e-services aim to identify a new scope of mobile e-services mainly addressing dynamic collected efforts between mobile users (enabled by M-P2P) and dynamic interactions with ambient environments (envisioned by LBS).

Since LBS was categorized into four major types (transaction services, information services, navigation and tracking services and safety services [3]), we accordingly exemplify some scenarios reflecting the LBS types as well as exhibiting dynamic collected efforts based on M-P2P.

We introduce an ambient shopping mall scenario to be a representative of ambient transaction services. The shopping mall scenario can easily be modified to act as ambient safety services. The information items distribution cooperation scenario then is considered as an exemplar of ambient information services. The cooperative peer group scenario then can be instantiated into ambient navigation and tracking services. As to the ambient location information acquisition scenario, it can be considered as an instance of ambient transaction services, ambient information services or ambient navigation and tracking services.

These applications differ from the past mobile e-services in addressing dynamic collected efforts between mobile users and dynamic interactions with ambient environments. The collective efforts of mobile users are not possible in the primal mobile services framework that deploys the services with the client/server architecture. Moreover, mobile devices in ambient e-services applications are very personal to the users, and thus the social contexts retained in the personal devices (e.g., the social relationships in the vicinity) can unfold the e-services of high complexity and security than that of no social contexts in sensor networks.

E-services with the social context environment render ambient e-services capable of providing the most appropriate collaborative power for mobile users anywhere whenever needed.

In the following subsections, each scenario will be detailed in terms of its motivation and promised value and the interaction structure.

3.1 Ambient Shopping Mall Scenario

In Figure 3, a picturesque view of the ambient shopping mall scenario is presented. In a shopping mall (fully equipped with wireless network infrastructures), information items (e.g., advertisement or sales promotion information) can be broadcast to passing-by peers with Info_BC_Station. This means the attained information items of a peer vary based on the peer's interactions with the shopping mall ambient environments.

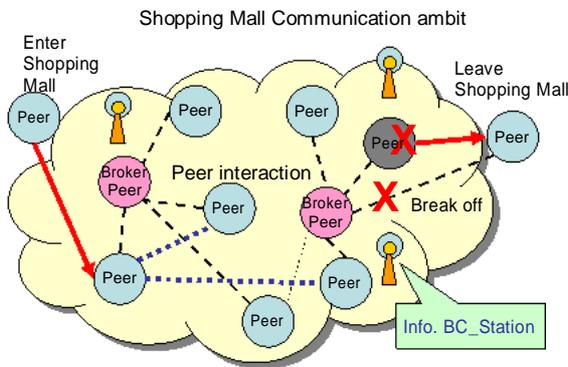


Figure 3. Ambient shopping mail scenario

The ambient e-service addressed in this scenario will be delineated for the case of new customers of high buyer perishability (entering the shopping mall and being in a rush to buy certain items without the knowledge of where to buy and how to buy cheap given relevant sales promotion). Those new customer peers carrying mobile devices can communicate with nearby peers when they just enter the shopping mall.

In such scenario, the goal of ambient shopping mall e-service is to provide a new channel for customers to get needed information. The assumption of this scenario is that the contents of information items kept in peers are different from one another based on their membership status. Mobile users will receive the information items based on their user context such as their preferences or membership data. Those members with higher loyalty points will get a higher discount e-coupon. This would encourage customers to join the membership program.

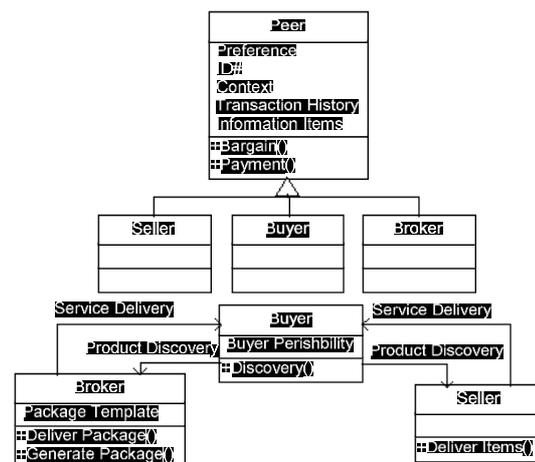
All mobile peers have their temporary peer ID number, preference, context data, transaction history and the information attained information items. Peers can do basic operations such as bargain and payment. Mobile peers can apply different role modules to do specific operations. There are the buyer module, the seller module and the broker module for mobile peers.

When a peer applies the seller module, it can respond a buyer's product discovery operation. If a seller peer has the information items that the buyer required, the seller then responds the buyer's request and starts a bargain operation with the buyer peer. After the payment process, the seller delivers the information items to the buyer. The operations performed by the buyer module are complementary to those performed by the seller module. But the buyer module has been designed not only to handle the transactions with seller module, but also to request the package service from the broker module.

The broker module is applied to those peers who will stay in the mall for a long time and own much complete information items. The broker module is designed for peers serving as service providers. The broker module has the 'generate' package, which can generate services of producing information items of the buyer's interests and preference. Package templates are used for the broker module to produce a service package in an easy way. Also, the broker module has the same ability as the seller module that can deliver the products.

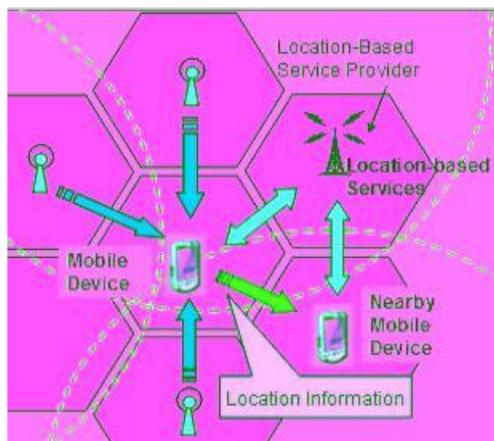
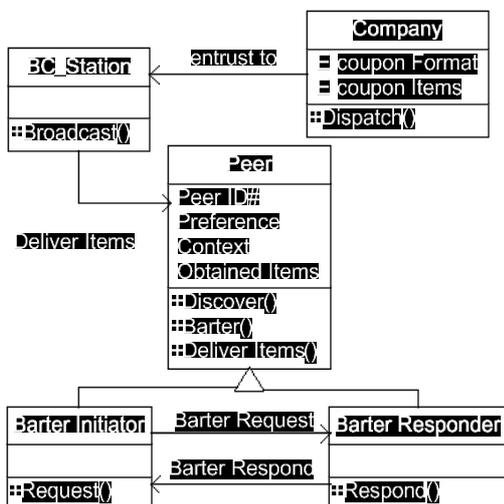
Although the effective reach is one of the shopping mall's goals. If the contents of information items are the same for everyone, both client/server and P2P designs can pass the items to the public but in different ways. However, mass broadcast like spam mails will annoy customers and thus be reluctant to join this shopping program, attaining the contrary results. When considering user differentiation, content differentiation will be in relation to user profiles and to ambient context sensors in the environment stack (that subsequently will involve social context and correlate to the privacy concerns of the customers). Those concerns make the P2P design is more applicable than the client/server architecture.

A UML class diagram is depicted (as shown in Figure 4) represents the relationships between the different modules as well as the peer module basic functions.



ambient e-service based on a mobile advertising infrastructure. Distributor companies can broadcast shops' advertisement items (e.g., e-coupons) through an ambient advertising infrastructure, which dispatches advertisement items relevance to their preferences to proximate users. In this scenario, e-coupons are considered as valuable information items. When a mobile user passes through a shop and wants to buy certain products but is short of the relevant e-coupons, the mobile peer can probe the surrounding environment to see if there are some nearby peers that own the desired e-coupons.

The P2P design combined with social awareness will make the connections not only in a rapid growth but also with persuasiveness, being more powerful than the client/server design. Moreover, as mentioned in Section 3.1 e-coupons become a valuable information item if the contents of information items are different. In this business model, mobile users can exchange their e-coupons, get some financial benefit, and be attracted to join the program so as to create a large scale of the participants. That is a win-win situation is created for both customers and the shops.



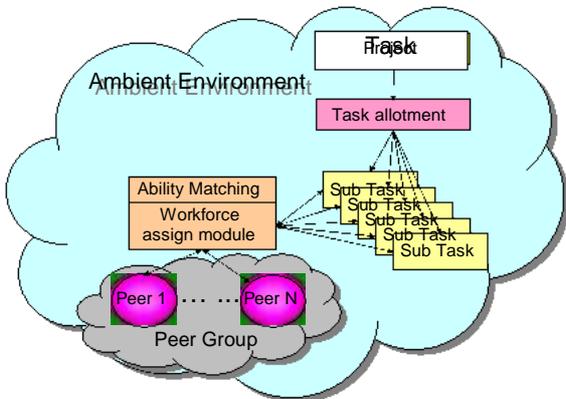
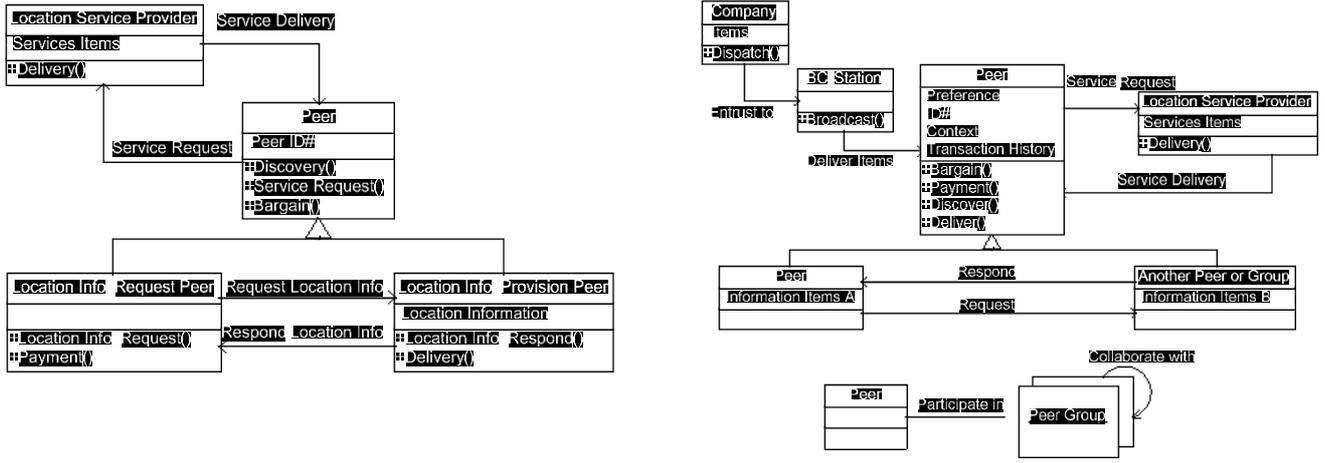
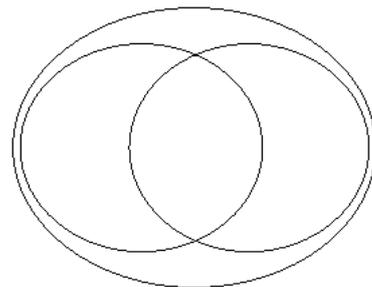


Figure 8. Cooperative peer group scenario

3.5 Ambient e-services Infrastructure



4.1 Trust, Reputation and Privacy Issues

Ambient e-services rest the focus on the interactions and collective efforts of the surrounding environments and nearby participating peers. However, in the ambient environments of ad-hoc networks, peers are not familiar with others. Not like the fixed Internet environment, there are no permanent databases of historical data that can be analyzed in the ambient environments. Accordingly, there exist no mechanisms for peers to trust each other in the ad-hoc network domain. However, before cooperating with other peers, peers need to guard themselves and seek who can be trusted based on the peers' reputations.

Privacy may have different meanings and significance to different mobile users. Privacy must be considered together with all other requirements-functionality, usability, performance, costs, security, and so on [6]. The privacy issue is not only important in the fixed Internet environment, but is even more important in the ambient environments.

The privacy concern is the barriers for mobile users join ambient e-services. When users are not feel safety for their sensitive data, they may not participant in such e-services. However, the more participants, the power of ambient e-services will provide. The social effort appears only when the number of the connections reaches above a certain threshold. This is the central problem for developing ambient e-services. Therefore, it is interesting to identify the privacy requirements for ambient e-services. Having decided the privacy issue for ambient e-services application, the interaction between reputation data and trust mechanism become meaningful. The semantic interactions of privacy, reputation and trust are depicted in Figure 11.

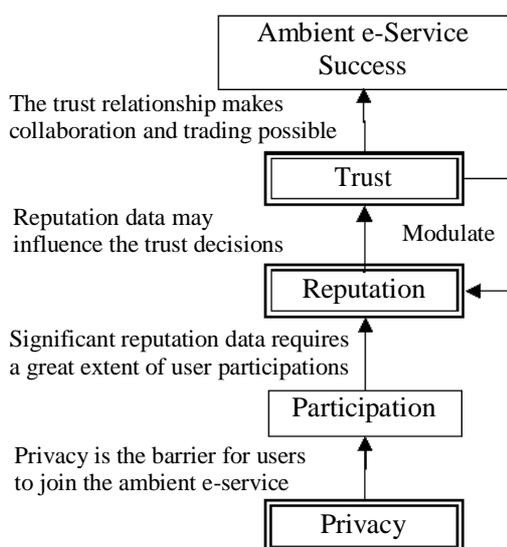


Figure 11. Semantic interactions of Privacy, Reputation and Trust

Many distributed trust management systems have been studied [5, 7, 9]. In a distributed trust system, reputation

information is scattered among parties in the system. For instance, a distributed trust and reputation management framework for e-services was proposed to solve the trust issues in a distributed environment, a distributed broker-based architecture which gather enough information to make a confident judgment of the trustworthiness [5]. Nevertheless, this solution is so designed with the client/server architecture that the brokers and the users live with the fixed Internet. Although this system like the ambient framework, use the collective efforts of the broker peers to build up the trust and reputation management system. However, this may work in the fixed Internet; but it is not suitable for the ambient e-services environments. The rationale is two-fold.

First, peers in ambient environment form a dynamic organization structure. Not like in the fixed Internet environment, peers do not have permanent ID in the ambient environments. A peer enters an ambient environment ambit and gets a temporary peer ID with the technology of mobile identity management [10]. Lack of permanent identification ID, the trust and reputation data cannot be tracked as the framework mentioned. Accordingly, unlike the fixed Internet environment the trust and reputation information cannot be retained. Second, an ambient environment's setting is different form the fixed Internet. In the fixed Internet environment, all brokers can communicate with each other. But in ambient environments, due to peer connections are in a limited range, peers may only connect to the surrounding peers. That makes the framework not applicable to the ambient e-services environments.

In short, there exist distributed trust and reputation systems to provide methods for collecting participators' reputation information for e-services. Unfortunately, it does to suit for the ambient e-services environments due to the difference in environment architectures and design ideas. These challenges are still awaiting future research to counteract.

4.2 Social Issues and Economic Issues

With the advent development of trust, reputation and privacy control for safety of ambient e-services, the relevant social issues and economic issues can be subsequently investigated.

The ambient e-services will develop certain revolutionary business models, which definitely needs to be paid attention in this new ambient era. User interface and the usability of the ambient e-services are quite different from the primal mobile commerce that should also be concerned as a vital social issue awaiting future studies. Not only technical problems being barriers to ambient e-services implementation, the legal issues are also troublesome. How to protect intellectual property in the ambient e-services [11]? How to treat the ambient e-service taxation?

Regarding the economics issues, one of the most important issues is the pricing utility issue. The ambient e-services require several kinds of supporting infrastructures, such as sensor networks, wired/wireless network infrastructures, broadcast stations and the e-service application providers.

The ambient value is delivered with the combined efforts of the supporting infrastructures. Accordingly, how to price ambient e-services and how to divide the revenue between the infrastructure providers will become an important issue.

In certain ambient e-service scenarios, the pricing issue is regarded as the major problem. For example, in the ambient shopping mall scenario, a buyer peer requests information packages from a broker peer. How the broker peer price the service package for different buyer peers is an important issue. Based on the buyer's anxiousness and preference, the broker can provide the buyer several packages of services in different prices. The buyer can also bargain with different brokers before concluding a transaction. In the information items distribution cooperation scenario, mobile users may exchange their information items using the barter mechanism. When bartering with other peers, how to evaluate the value of certain information items and how to make a fair-trade agreement are also belonging to the pricing utility issue. Even for the ambient location information acquisition scenario, the mobile users who own the location information can be treated as sellers holding information valuable to nearby mobile users (who want to engage location-based services but do not have the ability to know the location information). Mobile users with the location information can also price the information items based on buyer's preference and anxiousness.

Besides the pricing issue for the ambient interactions, the issue also occurs to the network infrastructure utility rates rendered for ambient e-services providers.

In cases that the infrastructures are built by governments, different networking enterprises and different service providers, different vertical or horizontal relationships result in different pricing and revenue sharing strategies. In short, how to price the e-services and how to divide the revenues of e-services will unfold as an important problem when the ambient e-services industry forms. Accordingly, the pricing utility issue of ambient e-services is a major economic issue worthy of future intense research.

5. CONCLUSIONS AND FURTHER RESEARCH

In this paper, the notion of ambient e-services is defined to identify a new scope of mobile e-services, addressing dynamic collected efforts between mobile users (enabled by M-P2P), dynamic interactions with ambient environments (envisioned by LBS), and the moment of value (empowered by wireless technologies).

Ambient e-Services applications can be divided into two types. One is for the distributed trading; another is for the distributed collaboration. However, social context and significant rapid growth of connections enable by P2P are the two major incentives for applying ambient e-service to such revolutionary business models.

The paper also presents an ambient e-services framework characterizing three supporting stacks. The ambient value stack describes the value process in ambient environments. The ambient technology stack identifies the technology

process to ensure connectivity and security in ambient interactions and cooperation between peers and then realize powerful collective efforts. The environment stack then represents the ambient basics for the collaborations.

We also exemplify several ambient e-service applications. Those applications differ from existing mobile e-services (grounded on client/server design) in terms of the focus of the dynamic interactions between peers in dynamic ambient e-service environments.

Finally, we identify a few future important research issues as follows: (a) trust, reputation and privacy control in ambient e-services environments, (b) pricing utility issues such as how to price the e-services and how to divide the revenues between the ambient e-services infrastructure providers (c) new business model in ambient e-service environments, (d) the taxation problem for ambient e-services (e) the intellectual property rights issues. Many of the traditional issues (such as QoS, fault tolerance, etc.) are also need to be revisited now that suitable solutions for the ambient e-service environments are desired.

REFERENCES

- [1] Varshney, U., R. Vetter, "Mobile Commerce: Framework, Applications and Networking Support", *Mobile Networks and Applications*, Vol.7, pp185-198, 2002
- [2] Kant, K., Iyer, R., Tewari, V. "A Framework for Classifying Peer-to-Peer Technologies". *Proceedings of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID'02)*, 2002
- [3] Lim, E., Saiu, K., *Advances in Mobile Commerce Technologies*, Idea group publishing, 2003
- [4] Katz-Stone, A., "Wireless revenue: Ads can work [online]", <http://www.wirelessauthority.com.au/r/article/jsp/sid/445080>, 2001
- [5] Lin, K., Lu, H., Yu, T., "A Distributed Trust and Reputation Management Framework for E-Services", *IEEE International Conference on Services Computing*, Shanghai, China, 2004.
- [6] Yu, E., Cysneiros, L.M. "Designing for Privacy in a Multi-Agent World " In: R. Falcone, S. Barber, L. Korba and M. Singh (Eds) "Trust, Reputation and Security: Theories and Practice": Springer-Verlag, LNAI 2631, pp209-223, May. 2003
- [7] Xiong, L. Liu, L. "Peer Trust: Supporting Reputation-based Trust for Peer-to-Peer Electronic Communities", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, No. 7, pp843-857, July, 2004,
- [8] Lindwer, M., Marculescu, D., Basten, T., Zimmermann, R. Marculescu, R., Jung, S., Cantatore, E. "Ambient Intelligence Visions and Achievements: Linking Abstract Ideas to Real-World Concepts," *Proc. Design, Automation and Test in Europe Conf.*, Munich, Germany, March, 2003
- [9] Yu, B., Singh, M. "A Social Mechanism of Reputation Management in Electronic Communities", *Proceedings*

- of the 4th Int. Workshop on Cooperative Information Agents*, pp154-165, 2000
- [10] Roussos, G., Peterson, D., Patel, U. "Mobile Identity Management: An Enacted View", *Int. Journal of E-Commerce*, Vol.8, No.1, pp81-100, 2003
- [11] Pitkänen, O., Mäntylä, M., Välimäki, M., Kemppinen, J. "Assessing Legal Challenges on the Mobile Internet" , *International Journal of Electronic Commerce*, Vol. 8, No. 1, pp101-120, Fall, 2003
- [12] Schilit, B., Adams, N., Want, R. "Context-aware computing applications", *Proceedings of the 1st International Workshop on Mobile Computing Systems and Applications*, pp85-90, 1994

An Extended Collaborative Filtering-based Recommendation Procedure for Multimedia Contents in M-Commerce

Meeyeon Kang¹, Yoonho Cho², Jaekyeong Kim¹

¹ School of business Administration, Kyung Hee University, Seoul, 130-701, Korea

² School of e-Business, Kookmin University, Seoul, 136-702, Korea
kkangmee@lycos.co.kr, www4u@kookmin.ac.kr

ABSTRACT

As mobile market grows more and more fast, the mobile contents market, especially music contents for mobile phones have record remarkable growth. In spite of this rapid growth, mobile web users experience high levels of frustration to search the desired music. And new musics are very profitable to the content providers, but the existing CF system can't recommend them. To solve this problem, we propose an extended CF system to reflect the user's real preference by representing users in the feature space. We represent the musics using the music's content based acoustic feature like timbral, MFCCs, rhythmic, and pitch contents in multi-dimensional feature space, and then select a neighborhood with distance based function. And for new music recommendation, we match the new music with other users' preference. To verify the performance of the proposed system, the simulation imitating the real user's decision-making and context in conducted. Through comparison with the pure CF, we validate our system's performance

Keywords: collaborative filtering, recommender system, mobile commerce, music information retrieval

1. INTRODUCTION

With the popularity of the mobile device, technologies and applications, the mobile contents market grows very rapidly. In mobile web environment the multimedia contents like musics, character images, and pictures are the most popular contents, and especially the music contents take a very large portion among them [7]. Although the popularity of the multimedia contents increases very fast, many users frustrate to search the music which they really want because of mobile phone's limitations like the small LCD, tiny keypad, and low memory and music's unique character. Music is hard to explain its feature like tempo, pitch, beat and other spectral features; therefore many users have to search for the desired music by scanning the offered music list one by one or keyword directly. For reducing mobile users' searching efforts and time, contents providers would adopt a recommender system.

Collaborative filtering (CF) is known as the most successful system in recommender systems. But the CF system results in many problems because of its input data representation. The general CF system uses a user-item binary matrix; therefore it is hard to find the real neighbors.

And in mobile environment new musics are very frequently supplied, and their purchasing ratio is considerably high, however existing CF systems can't recommend new musics. A new music doesn't be purchased by anyone, it can't have any ratings.

In this paper we propose the system which can solve these problems. This system extracts the music's various content-based acoustic features like MFCC, tempo, beat, pitch, and then represents these musics in a

multi-dimensional feature space. As doing this, we can identify the neighbors who are really similar to the target. And we can recommend new musics by giving a virtual rating to them in the feature space. This paper describes how can solve these problems in mobile web environment.

The rest of the paper is organized as follows: Section 2 reviews the related work on CF system and music information retrieval. Section 3 gives a brief overview of the proposed system. Section 4 presents the detail recommendation procedure. Section 5 describes the experimental design and evaluations for system. Finally, Section 6 provides our conclusions.

2. RELATED WORK

2.1 Collaborative Filtering

Collaborative filtering (CF) is the most familiar, most widely implemented and most mature of the technologies in recommender system [1]. It is defined as one which makes recommendations by finding correlations among users of a recommender system [8]. The goal of CF system is to suggest items for a target user based on the user's previous preference and the opinions of other like-minded users. But the CF system has many problems like sparsity, and scalability problem, but these problems have been explored in [2,6,10,11].

Beyond these problems the CF system has a radical problem related its input data representation. The CF system uses a user profile which is composed with m user's ratings about n items, and is represented by the $m \times n$ user-item binary matrix. When a user has purchased a music, a general CF system gives a rating to the music. This causes that the system is hard to identify

user's real neighbors because it uses only common musics for forming the target user's neighborhood. Therefore it is possible that someone whose preference is similar to the target user's preference didn't purchase common musics.

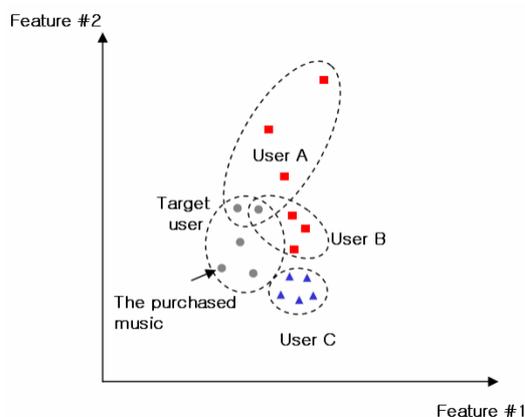


Figure 1. CF system in feature space

Figure 1 shows an existing CF system's problems. Generally the existing CF system selects user A as the target user's neighbor because he purchased the common music. But user A's preference is quite different from target user's preference when their preferred musics represent in feature space. In Figure 1, user B and user C are more similar to the target user rather than user A. Moreover the system can't find hidden neighbors such user C as the target user's neighbor.

And the CF system has the new-item ramp up problem. The new item ramp up problem is the most difficult and unsolved problem in the CF system. In the mobile web environment, new musics are very frequently provided, their purchasing ratio is very high, and many contents providers try to find out this problem. But the existing CF system can't recommend the new music.

In Section 4 we describe how to solve the problems related the input data representation and new item ramp-up problem.

2.2 Music Information Retrieval

As the popularity of digital music has been increasing, efficient and accurate automatic music information processing will be an extremely important issue, and many systems try to classify the music genre, and describe the music contents automatically. Many researches have suggested the automatic music classification methods, especially music feature representation. Many different features can be used for music classification, e.g., content-based acoustic features including tonality, pitch, and beat.

The content-based acoustic features are classified into timbral features, rhythmic content features, and pitch content features. Timbral features are mostly originated

from traditional speech recognition techniques. Typical timbral features include Spectral Centroid, Spectral Rolloff, Spectral Flux, Energy, Zero Crossing, Mel-Frequency Cepstral Coefficients (MFCCs) [9].

Rhythmic content features constrain information about the regularity of the rhythm, the beat and tempo information [9]. Pitch content features deals with the frequency information of the music bands and are obtained using various pitch detection techniques.

3. OVERALL PROCESS

We design to deal with the problems of the CF system's input data representation and a new music recommendation on the mobile web. The system follows a general CF process, and is added a new music recommendation module.

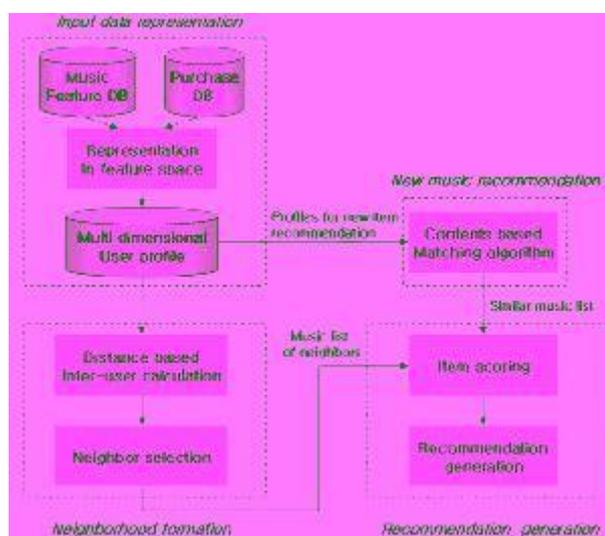


Figure 2. The overall process

The input data representation module creates a user profile using the music feature database and purchase database. The Music feature database is organized with the features like timbral, MFCC, rhythmic, and pitch features, which are extracted by the MARSYAS system [12]. The Purchase database is organized with the individual user's purchasing or pre-listening histories. Using this module, the system reflects the individual user's preference, can represent each user as a point in the multi-dimensional feature space.

The neighborhood formation module selects the target user's neighbors. The existing CF systems calculate the inter-user correlation by the Cosine formulation or Pearson Coefficient; however this way is hard to identify the neighbors who are really close to the target user as described Section 2.1. Therefore this module calculates the distance between the target user and other users to identify his real neighbors, and forms his neighborhood. Finally it gathers neighbors' purchasing or pre-listening musics, and gives the list of these musics to the recommendation generation module.

The recommendation generation module creates a recommendation list for the target user. This module scores the musics which neighbors purchased or pre-listened. As many neighbors purchase or pre-listen, this score increase.

The new music recommendation module solves the new item ramp-up. This module finds the most similar music by the content-based matching algorithm, and generates a similar music list, and gives it to the recommendation generation module.

4. RECOMMENDATION PROCEDURE

4.1 User Profile Creation

A user profile, which is the key component for CF system, includes information about user's preferences on music. This module is organized by two phases, the first phase is the extraction music's features, and second phase is the construction of user's preferred music set.

Phase 1: The extraction music's features

For extracting the music's features, we use the described features in section 2.2. For getting these features, we use the MARSYAS system [10]. This system extracts the timbral, MFCCs, rhythmic, and pitch content features. The timbral features consist of average spectral centroid, average spectral rolloff, average spectral flux, energy, and zerocrossing. MFCC features consist of the mean or each of the first five MFCC coefficients over the frames. The rhythmic features consist of six features from the rhythm histogram. Pitch features are five features from the pitch histograms.

The system extracts the whole music's features, represents the individual musics in the multi-dimensional feature space as a point.

Phase 2: The construction of user's preferred music set

The general CF system uses the user-item binary matrix, but we use the user's preferred music set as the user profile. In the mobile environment users could do actions like purchasing, pre-listening or giving up buying a music. We set a rating about user's action assuming that these actions reflect user's preference about the music. Therefore we define a rating about user's action.

If user 'a' has purchased music 'x', it could be given a rating ' R_a^x '. We define the ' R_a^x ' as the following:

$$R_a^x = \begin{cases} 1.0 & \text{If user 'a' has purchased music x} \\ 0.5 & \text{If user 'a' has pre-listened music x} \\ 0 & \text{If user 'a' has done any action to music x} \end{cases} \quad (1)$$

With the music which user 'a' has done any actions to and its ratings, the system composes the individual user's

preferred music set. We defined user 'a's preferred music set as the following:

$$\text{Preferred music set} = \{(x_1, R_a^{x_1}), \dots, (x_{N_a}, R_a^{x_{N_a}})\} \quad (2)$$

N_a is the number of the music whose R_a^x is non zero, and x_i is the music which user 'a' purchased or pre-listened.

The illustration of the preferred music set is showed in Figure 3.

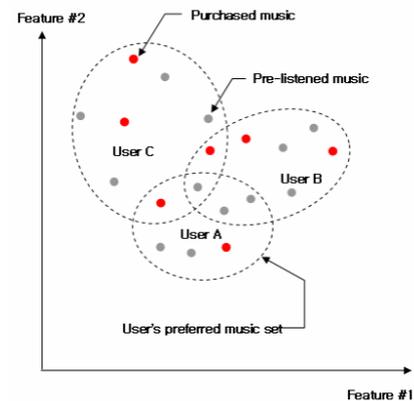


Figure 3. User's preferred music set

In Figure 3, user 'A', 'B', and 'C' are represented in the 2-dimensional feature space. The system represents all of the musics in this space, and binds the purchased and pre-listened musics as user's preferred music set by a dotted line. As shown in Figure 3, the user A's preferred music set includes 7 musics, and consists of 2 purchased musics and 5 pre-listened musics, which reflect his preference. User B and C's preferred music set also are composed by this way.

As doing this way, the user profile is composed by the whole users' preferred music sets.

4.2 Neighborhood Formation

Because whole users and musics are represented in multi-dimensional feature space, the system can find the target user's neighbors with distance based inter-user similarity calculation. If a user is close to the target user in this space, he could have a similar preference the target user's one. To calculate the distance between users, we assume that a user's preferred music set is a cluster.

There are many inter-cluster distance function, and the favorite function is the Centroid Euclidean distance function. This function is very simple and easy to calculate, but it works well in a condition that each cluster is fairly distributed and its form is a circle. However the individual users' preference isn't same and their distributions are difference from each other in shown Figure 3. Therefore we use the Average

Inter-Cluster distance function [3] which the distance function considering the variance of two clusters, and we modify this function a little as the following:

$$d(a,b) = \frac{\sum_{y=1}^{N_a} \sum_{x=1}^{N_b} \sqrt{\frac{\sum_{i=1}^N (R_a^x \cdot F_x^i - R_b^y \cdot F_y^i)^2}{N}}}{N_a \cdot N_b} \quad (3)$$

Where $d(a, b)$ is the distance function between user a and b , N is the number of features. N_a and N_b is the number of user a 's and user b 's preferred musics, respectively. R_a^x and R_b^y is the rating about music x by user ' a ' and about music y by user ' b ', respectively. F_x^i and F_y^i is the value of i th feature of music x and music ' y ', respectively.

We sort the neighbors ascendingly according to the distance value, and then select the L neighbors. Finally we determine the neighborhood set H for target user c , and $H = \{h_1, h_2, \dots, h_L\}$ such that $c \notin H$.

4.3 Recommendation Generation

The proposed system generates a list of n musics, $R = \{r_1, r_2, \dots, r_n\}$ such that $r_i \notin \{\text{the music that the target user } c \text{ has already purchased}\}$. $PLS(c,x)$ denotes the Purchase Likelihood Score of the target customer c on music x , and is computed as the following:

$$PLS(c,x) = \frac{\sum_{a \in H} (R_a^x - \bar{R}_a) \times sim(c,a)}{\sum_{a \in H} sim(c,a)} \quad (4)$$

Where \bar{R}_a is the user a 's average rating, u, w are the other users included the neighborhood set H . Because we use the distance function, we have to covert these distance values into similarity values, and normalize them. $sim(c,a)$ function denotes the similarity between target user c and user a , and it is computed as the following:

$$sim(c,a) = \frac{Max_{u,w \in H} [d(u,w)] - d(c,a)}{Max_{u,w \in H} [d(u,w)] - Min_{u,w \in H} [d(u,w)]} \quad (5)$$

4.4 New Music Recommendation

In mobile web environment, new musics are frequently provided and their purchasing ratio is very high, but the existing CF system can't recommend new musics. The general CF systems use the common items to find neighbors. Therefore they can't recommend musics that anyone didn't have been purchased. New musics can't be purchased by anyone, therefore they don't have ratings. If a new music has a rating, it could be recommended. To solve this problem we give a virtual rating to the new music.

Phase 1: Determination of new musics recommendation

New music also has the above mentioned features, which can be extracted. The proposed system extract new musics' features by the same way described in Section 4.1, and represent the musics in multi-dimensional feature space. Although new music can't have any ratings, it could be included in any user's preferred music set. Therefore we define a condition to determine new music's recommendation, as the following:

$$\begin{aligned} & \text{If } x_{new} \in a \text{'s preferred music set} \\ & \text{then } x_{new} \text{ is recommended} \end{aligned} \quad (6)$$

Where x_{new} is a new music, $a \in H$, $H = \{c, h_1, h_2, \dots, h_L\}$, and preferred music set is defined in (2). Why target user c is included in the neighborhood set H is that the new item doesn't purchased by target user, and the system have to reflect the target user's preference as much as possible.

The target user's preferred music set reflects his preference, therefore a new music comes in the target user's preferred music set or his neighbor's one, it would be preferred by them. The illustration of this is showed in Figure 4.

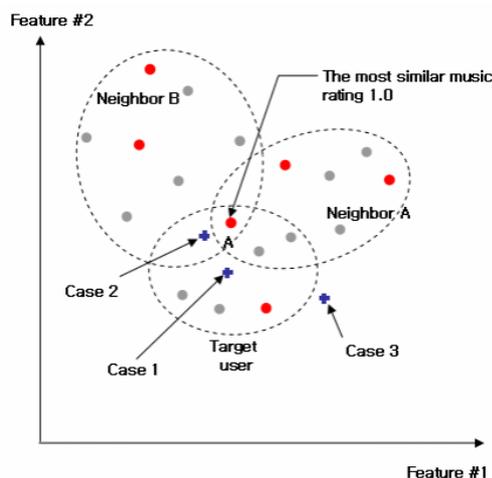


Figure 4. New music recommendation

Case 1, 2, and 3 are new musics and represented in the feature space. As shown Figure 4, case 1 and 2 can be recommended because they are satisfied the condition, but case 3 isn't. Therefore the system determines to recommend the case 1 and 2.

Phase 2: Finding the most similar music

We determine whether to recommend the new music or not, but it still doesn't have a rating. To give a virtual rating to the new music, the system searches for the most similar music where it is included. If music x and y in the same user's preferred music set are very similar, they would be given the very similar rating by that user. Most similar music is founded as the follows:

$$\min_{y \in a} \{d(x,y)\} = \sqrt{\frac{\sum_{i=1}^N (x_{new}^i - y^i)^2}{N}} \quad (7)$$

Where x_{new}^i and y^i is the i th feature value of new music x_{new} and music y included user a 's preferred music set, respectively.

The illustration of this is showed in Figure 4. As shown in phase 1, the case 1 and 2 can be recommended, but we see the case 2. The case 2's most similar music is the music A, and its rating value is 1.0. Therefore the case 2 is given the rating 1.0.

Through this way new music can have a rating, with this rating the system calculates new item's PLS value. Therefore we generate a recommendation list with combination of old musics and new musics.

5. EXPERIMENT EVALUATION

5.1 Experiment Design

For the purpose of the performance evaluation of the proposed system, we developed a simulation program imitating the real user's decision-making and context in conducted. And we use two databases for our experiment. The first dataset contains features of 500 songs which are the most popular songs in mobile environment. The second dataset contains 1000 users' utility. We assume that the individual user's preference is different from each other, and they have their own utilities. Therefore we create utility functions for 1000 users.

Because the CF system has to start with a few transactions, we make the initial transactions. This is conducted by user's utility, and transactions organized by three purchase and five pre-listened transactions. With these transactions, we conduct the initial experiment.

For the new music recommendation, we offer the 100 new musics to the system, conduct the second experiment. We give the new music with no ratings to the system, evaluate the system's performance.

5.2 Evaluation

To evaluate our system, we use two matrix widely used in the information retrieval community namely recall and precision. Each matrix is defined as the following:

$$\text{Recall} = \frac{\text{Purchased musics}}{\text{all recommended musics}}$$

$$\text{Precision} = \frac{\text{Recommended musics}}{\text{all purchased musics}}$$

However these two measures is often conflicting in nature. Then we use the standard F1 metric that gives equal weight to them both and is computed as the following:

$$F1 = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}$$

We compute F1 for each individual user and calculate the average value to user as our matrix.

6. CONCLUSION

Because of an information overflow for multimedia contents like music, character images, and movie, mobile users need recommendation systems. Especially, music is the most popular contents in the mobile contents market, but there are few music recommender systems to aid mobile users to search musics which they really want. The CF system is known to be useful, but it is inadequate for the multimedia contents. Moreover CF system can't recommend the new item which is very frequently supplied and most contents providers want to sell out. We proposed the extended CF system which can solve these problems in the feature space and verify our system's performance.

REFERENCES

- [1] Burke R., "Hybrid recommender systems: Survey and experiments", *User Modeling and User-Adapted Interaction*, 2002.
- [2] Cho Y.H., J.K. Kim, "Application of web usage mining and product taxonomy to collaborative recommendation in e-commerce", *Expert systems with application*, 2004.
- [3] Everitt B.S., S. Landau, M. Leese, "Cluster Analysis", *Oxford University Press*, 2001
- [4] Goto M., Y. Muraoka, "A beat tracking system for acoustic signals of music", *ACM mulmimedia*, pages 365-372, 1994.
- [5] Kim, C.Y., J.K. Lee, Y.H. Cho, D.H. Kim, "VISCORS: a Visual Contents Recommender System on the Mobile Web", *IEEE Intelligent Systems, Special issue on Mining the Web for Actionable Knowledge*, Accepted for publication, 2004.
- [6] Kim J.K, Y.H. Cho, "Using Web Usage Mining and SVD to Improve E-commerce Recommendation Quality", *LNCS (Lecture Notes Computer Science): Intelligent Agents and Multi-Agent Systems*, pp 86-97, 2003.
- [7] Korea IT industry Promotion Agency, "The Digital Contents Market Report", 2003.
- [8] Pazzani M.J., "A framework for collaborative, Content-Based and Demographic Filtering", *Artificial Intelligence Review*, 1999.
- [9] Rabiner L., B.H. Juang, "Fundamentals of Speech recognition", *Prentice-Hall*, NJ, 1993.
- [10] Sarwar B.M., G. Karypis, J.A. Konstan, J.T. Riedl, "Application of dimensionality reduction in recommender system - a case study", *ACM*, 2001.
- [11] Sarwar B.M., G. Karypis, J.A. Konstan, J.T. Riedl, "Analysis of Recommendation Algorithms for E-Commerce", *In Proceedings of ACM Conference on Electronic Commerce*, 2000.
- [12] Tzanetakis G., "MARSYAS: A framework for audio analysis", *Organized sound*, Vol.4(3), 2000.

An Overview of the Impact of National Culture on the Adoption of Mobile Commerce

Nasrin Rahmati

School of Business Systems, Faculty of Information Technology, Monash University, Wellington Road, Clayton, Victoria 3800, Australia

ABSTRACT

The predictions that mobile technologies would greatly benefit both firms and individuals have now come under increasing scrutiny. Some of the authors suggest that the industry must move beyond “nice-to-have” services and devise new “must-have” services that positively affect people’s lives (Jarvenpaa, et al. 2003). Others believe that mobile commerce is facing many obstacles as an emerging market, this is particularly so in some countries such as United States (Venkatesh et al, 2003). This study provides a brief overview of the literature on the impact of national culture on the adoption of mobile commerce. The study focuses on Singapore, Japan, UK and Germany.

Keywords: Mobile Commerce, Mobile payment, National Culture

1. INTRODUCTION

This study follows the suggestion made by Malone (2001). He suggests that academic Information Systems researchers accelerate the rate at which businesses learn from each other's experiments. The IS researchers observe the experiments, gather and analyse data in systematic and interesting ways and disseminate the lessons to anyone who wants to listen. Mobile commerce is distributed computing that involves elements whose location changes in the course of computing (Samaras, 2002). The opportunities associated with mobile commerce are the ability to compute, to do commerce and access information from anywhere and at any time. Payment over mobile networks, while bearing some relation to fixed-line eCommerce, will be very different. Firstly, mobile devices are unlikely to support traditional online shopping, such as buying a book or piece of furniture, even if higher speed mobile networks are available.

The mobile environment is, however, ideally suited for payments for digital content, and for payments that use the device as a secure, mobile payment instrument. Unlike eCommerce, mobile commerce is inherently global given the popularity of mobile phones in every major market worldwide.

It has been suggested by many authors (eg, Khalifa and Cheng, 2002) that mobile commerce is the next big wave of business. It is also suggested that given the already high mobile phone penetration the rapid growth of mobile commerce is assured. It is generally agreed that mobile phones have offered great convenience, but they lack the screen space and bandwidth available in traditional desktop computers. Although several technologies such as WAP/WML, Palm.Net, and HDML, have been developed specifically for mobile devices, it is agreed that there is no successful and dominant technology for displaying Web pages over mobile devices (Urbaczewski, et al, 2002).

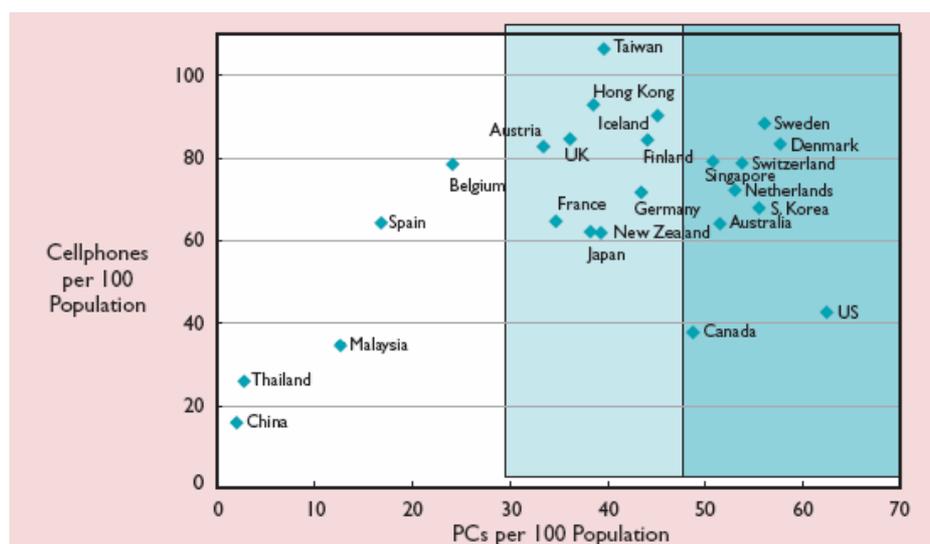


Figure 1: Cellphones and OCs in 2002 from “The Diffusions of Mobile Data Applications” by Elizabeth Fife and Francis Pereira

Figure 1 shows the difference between PC and Cellphones adoption in different countries. It is also suggested that although the use of mobile data applications has grown across the world, the adoption rates of these applications and services has been very uneven across countries. Figure 2 shows the different growth rates in adoption of PC and Cellphones by different countries.

Many of the standards designed to transfer information from one terminal to a mobile device (most notably with SMS, a fairly mature technology based on text messages of 160 characters or less), can also originate with a mobile device. Some of these technologies are

more developed than others, and some cultures and nations use these technologies more widely than others. Table 1 shows the relative penetration of SMS messaging in different nations (see Table 1).

It is important to capture the influence of the national environment, as contrary to early predictions, mobile commerce has not been adopted similarly by different nations. It has been suggested that those with well-developed infrastructures, resources and skills have gained the initial advantage (eg, Brown et al. 2004). In December 2000, Germans sent 1.8 billion SMS messages and Britons sent 803 million messages, both up about 50% from the levels listed in the table.

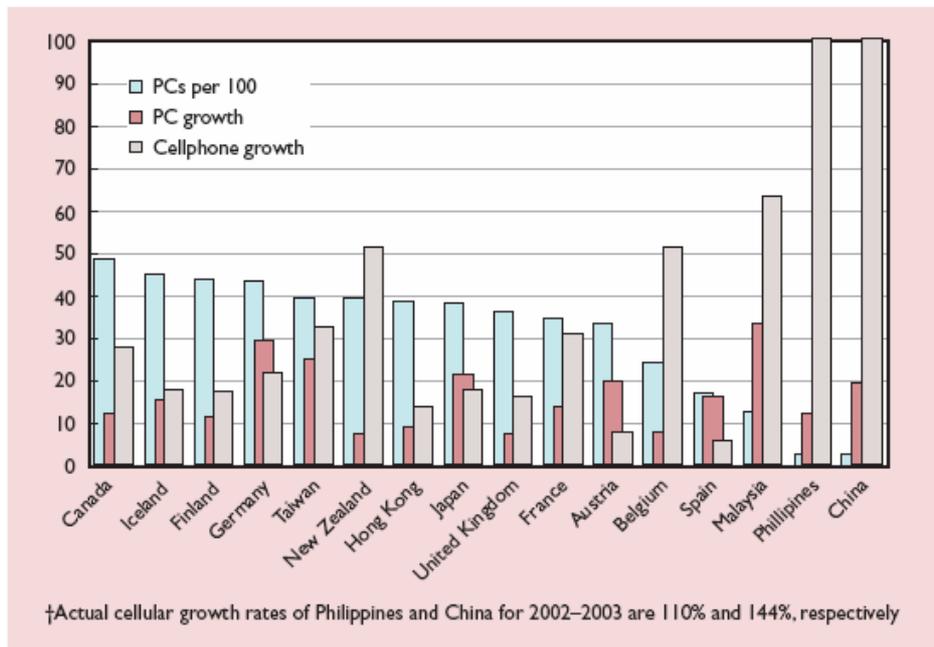


Figure 2: PCs, and growth rates of PC and Cellphones penetration (2000-2002) from “The Diffusions of Mobile Data Applications” by Elizabeth Fife and Francis Pereira

Nation	Population (millions) (June 2000)	Mobile Users (millions)	SMS sent per month (millions)	SMS/mobile user/ month
Australia	18.6	8.1	30 (3/2000)	3.7
Finland	5.2	3.4	175	51.47
Germany	82	28.2	1000	35.46
Norway	4.5	2.9	150	51.72
Philippines	73	4.1 (6/2000)	540 (12/2000)	131.7
Sweden	8.8	5.4	175	32.41
UK	59	28.3	400	14.13
USA	280	134 (7/2000)	20 (3/2000)	0.15

Table 1: Penetration of SMS in everyday life (Urbaczewski, et al. 2002)

SMS accounts for 12 percent of European mobile operators’ revenues. According to Forrester, 156 million SMS messages are sent in Europe every month. Nearly 90 percent of these messages are person-to-person (P2P) traffic. Application-to-person (A2P) traffic, such as weather reports sent by SMS, accounts for 11 percent of traffic. Forrester predicted SMS revenues to account for 47 percent of total messaging revenues in 2007. Multimedia messaging (MMS) will

account for 32 percent, Instant Messaging (IM), 10 percent, Mobile Email, 9 percent, and Enhanced Message Service (EMS), 3 percent. Forrester also forecasted that SMS traffic will rise to 11.5 billion messages per month in 2004. Still why is short message service (SMS) messaging so popular in Europe but barely used in the US? The use of SMS has been relatively low in the United States, as compared to Asia and Europe, where, in Europe alone, it is a \$14

billion dollar market. Of the estimated 576 million people reported to be using SMS, 256 million are in the Asia-Pacific Region and 196 million are in Europe. One of the crucial differences between European and

American companies, according to a survey conducted by Protogeris (2002) was the payment system for on-line sales.

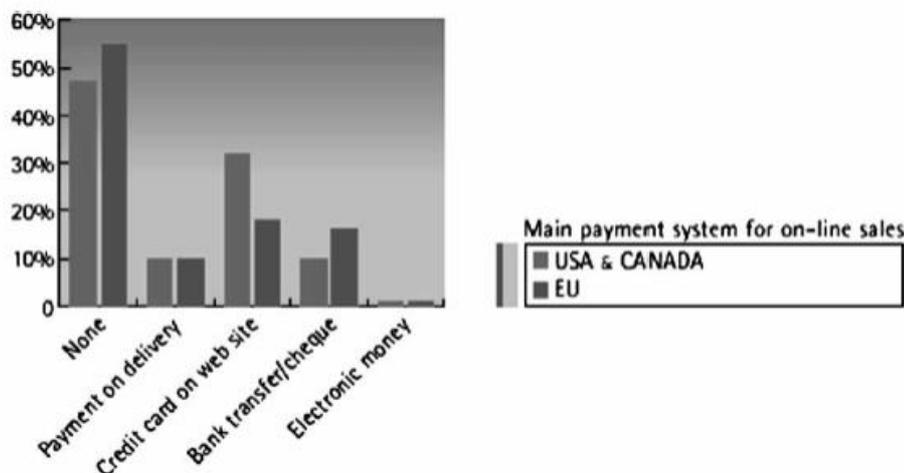


Figure 3: Payment systems for online sales a comparison between businesses from EU and US

Lee et al. (2002) used values as the core concept of culture in their comparative cultural study on the adoption of mobile commerce by Japan and Korea. The authors focused on values as the core concept of culture to address the impact of values on people behaviour and the attitudes which in turn affect the ways people behave in their lives (also cited in Geertz, 1973; Straub et al. 2002, Trompenaars and Hampden-Turner, 1993).

Xu et al. (2004) include different factors under Technology, Organizational and Environmental Contexts, to compare the ecommerce adopting countries with non-adopters. At technology level they are concerned with 1- technology readiness- which refers to technology infrastructure and IT human resources. 2- technology integration- which refers to the extent of inter-connection among IT systems and databases within and beyond firm boundaries. At the organizational level they are concerned with 1- firm size- which refers to the number of employees in the organization, 2- global scope and 3- managerial obstacles. Under the Environmental Context they look at competition intensity and regulatory environment. The present study examines Hofstede's work on national culture. Then using the culture level factors provided by his study compares the use of mobile commerce in Singapore, Japan, UK and Germany.

2. DIFFERENT APPROACHES TO CULTURAL STUDIES

There are different approaches to study the national culture impact on mobile commerce in different countries. Lee et al. (2002) on the basis of the findings of previous research on values (eg, Sweeney and Soutar, 2001, Sheth et al., 1991) suggest four dimensions of values. Their suggested dimensions are

Functional Values; Emotional Values; Social Values and Monetary Values. They use this four sub-value structure to interpret the different usage patterns of mobile internet users in Korea and Japan. Their findings indicate a significant difference between Korean and Japanese users in terms of the suggested dimensions of value structure.

Markus and Soh (2002) argue that although effects of national culture are an important line of research on global information management, the researchers should not lose sight of structural conditions related to ecommerce. They suggest structural conditions as physical, social and economic conditions that shape ecommerce activity.

Different models have been suggested for the adoption of any new technology, eg mobile commerce technologies. One of the suggested models is suggested by Xu et al (2004). To study the contextual factors that influence the innovation process they use earlier work by Tornatzky and Fleischer (1990) and suggest three aspects of a firm's context that influences the process by which it adopts and implements technological innovations: 1-Technological Context (both internal and external technologies with impact on the company). 2- Organizational context (eg, size, scope, managerial structure, etc). 3- Environmental context (eg, industry, competitors, government, etc). There are also suggestions regarding the combinations of many factors such as National Culture, Infrastructure and Access Costs on the adoption of mobile commerce in different parts of the world (Heales, 2004).

2.1 Hofstede's Framework On Culture

Hofstede identified five different variables in the study of culture, namely power distance, uncertainty avoidance, individualism, masculinity and long-term

orientation. Each of the five concepts of culture is defined as follows:

- Power distance: the degree in which members of a society accept that power is unfairly distributed (Teng, Calhoun, Cheon, Raeburn and Wong, 1999). Hofstede believes that High Power distance at work place means that subordinates expect to be told what to do (Hofstede, 1980).
- Uncertainty avoidance: the extent to which people in a society are unnerved by the unknown and ambiguity, thus leading them to look for conformity (Teng, et.al., 1999). Hofstede (1980) suggests that societies with high uncertainty avoidance need written or unwritten rules. They need formalization and standardization.
- Individualism: the degree in which a society believes that individuals should take care of themselves and their family as opposed to collectivism, whereby there is absolute loyalty towards a larger group (Teng, et.al., 1999). According to Hofstede (1980) in a collectivist society relationship prevails over task and values standards differ for in-group and out-groups.
- Masculinity: the extent in which a society adopts a competitive and assertive nature, constantly striving to achieve success over others as opposed to femininity where members of a society care for everyone and cherish relationships with others (Teng, et.al., 1999)
- Long-term orientation: the extent in which a society embraces tradition and values long-term commitments. A high long-term orientation society is said to be more resistant to change. A low long-term orientation society is one where traditions are not

generally observed and thus are more receptive to change (International Business Centre, 2003)

Heales (2004) only uses Uncertainty Avoidance, Performance Orientation (Javadian, 2001) and Future Orientation (Kluckhohn and Strodtbeck, 1961). Performance orientation is the degree to which a society encourages and rewards group members for performance improvement and excellence. Singapore, Hong Kong and New Zealand score the highest, Russia, Argentina and Greece the lowest, whereas Australia has a medium performance Orientation. Future Orientation is the extent to which a society encourages and rewards future-oriented behaviours such as planning and investing in future. Countries with high future orientation include Singapore, Switzerland and Netherlands.

3. A CROSS-CULTURAL REVIEW

Using Hofstede’s cultural framework above, four different countries will be examined to determine whether cultural differences in European and Asian countries will have an impact on the adoption of mobile payment applications. In this research, Germany, Britain, Singapore and Japan are used as case studies since mobile payments have been widely established in these countries. Tables will be used to summarise the various findings.

Judging from table 2 below, it appears that both regions although have different power distance indexes; both are receptive to mobile technologies.

JAPAN	SINGAPORE	GERMANY	BRITAIN
Medium power distance score = 54 Mobile phone used as a tool to reflect social status (Barnes & Huff, 2003) This can have beneficial impact on the adoption of mobile payment.	High power distance score = 74 Strong government policies imposed to ensure smooth and controlled adoption into the global networked economy (Warshauer 2001).	Low power distance score = 35 Mobile phone is not a status symbol (Hofvenschiold, 2003)	Low power distance score = 35 Mobile phone is not a status symbol (Hofvenschiold, 2003)

Table 2: Power Distance Between The East And West

However, it seems that there are some social pressure in adopting mobile technology in Asian countries, but this is absent in European countries. This is in agreement with the findings of Hofstede study. He

believed that in High power distance nations the adoption of the technology, if encouraged by the government or the superiors have a positive impact on the adoption by the members of the society.

JAPAN	SINGAPORE	GERMANY	BRITAIN
High uncertainty avoidance score= 92 Japanese people are used to having innovative entertainment devices being developed regularly. Thus there is not much uncertainty about Imode (Barnes & Huff, 2003)	Low uncertainty avoidance score= 8 The adoption process of mobile payment technologies is controlled and regulated by government organisation (IDA, 2003) Thus eliminating uncertainty.	High uncertainty avoidance score = 65 Mobile applications in Germany are popular, but not as much interest as Japan The reason being personal information security is a major concern for Germans (Martin, 2003)	Low uncertainty avoidance score= 35 More tolerant to innovative ideas, yet not capable of attaining full-scale implementations due to considerable demand for detail and timeliness (Hofstede, 1997) Mobile payment still not widely used as compared to Japan

Table 3: Uncertainty Avoidance between East and West

Based on table 4, Asian countries are more collectivist as compared to European countries. Societies with low individualism believe in 'learn how to do' and 'learning is for the young'. This can mean that the younger generation are expected to adopt the new technology faster than the older generation. On the

other hand in countries with high individualism, there is a belief about 'permanent education'. This combined with the 'self-interest' should serve as a positive impact on the adoption of the mobile commerce by different generations alike.

JAPAN	SINGAPORE	GERMANY	BRITAIN
Medium individualism score= 46 Main application of Imode is E-mail. (Mizukoshi, Okino & Tardy, 2001). Thus suggesting keeping in touch is important to the Japanese	Low individualism score= 20 Increased in work hours and mobility in society resulted in mobile phones being commonly used to keep in contact with family (Burgess, 2004)	High individualism score= 67 Paybox, a mobile payment application Imode is more popular than Imode. (Pousttchi & Zenker, 2003). Thus suggesting personal convenience is more important.	High individualism score= 89 Mobile entertainment and small payments like M-pay and M-live are most popular mobile applications (Shannon, 2003). Thus suggesting personal convenience is more important.

Table 4: Individualism between East and West

It appears that there are some conflicting findings into whether masculinity can affect the rate of mobile payment adoption in countries of different regions. This is because in Western countries where masculinity is high, people do not necessary use the mobile phone as a tool to achieve competitiveness. On the other

hand, people in Eastern countries such as Japan, a highly masculine society, tend to use the mobile phone to gain a competitive edge. Therefore, it should be noted that masculinity does not necessary influence the rate of adoption of mobile payments in western countries as it does to eastern countries.

JAPAN	SINGAPORE	GERMANY	BRITAIN
High masculinity score= 95 Need to read important e-mails and news anytime anywhere to be competitive (Fife & Pereira, 2003).	Medium masculinity score= 48 Mobile payment applications mainly used for making payments such as paying for car parking (IDA, 2003). Thus suggesting personal convenience than competitiveness	High masculinity score= 66 Relatively low interest for Imode in Germany as compared to Paybox, which is used for making mobile payments, attracted more users (Pousttchi & Zenker, 2003). Thus suggesting personal convenience than competitiveness	High masculinity score= 66 Mobile entertainment and small payments like M-pay and M-live are most popular mobile applications (Shannon, 2003). Thus suggesting personal convenience than competitiveness.

Table 5: Masculinity between East and West

JAPAN	SINGAPORE	GERMANY	BRITAIN
High long term orientation score= 80 Imode is locally developed and proved very popular. No foreign mobile commerce application in Japanese market as yet.	Medium long term orientation score= 48 A few locally developed mobile payment applications with foreign company collaborations Government regulates the rate of adoption (IDA, 2003)	Low long term orientation score= 31 A few locally developed and foreign mobile payment applications available on the German market.	Low long term orientation score= 25 A few locally developed and foreign mobile payment applications available on the British market

Table 6: Long term Orientation between East and West

As can be seen from table 5, it can be concluded that in eastern countries, although have differing long-term orientation indexes; the people are generally more receptive of locally developed mobile payment technologies which is developed over time, for example iMode. Western countries on the other hand, have a number of local and foreign mobile payment applications being successfully implemented into the market.

addition to examining the impact of national culture on the adoption of mobile commerce one should not lose sight of physical, social and economic arrangements that shape ecommerce business models and influence individual use of these technologies. The next step would be to develop a mobile commerce adoption model by individual users considering different factors including national culture, economic, social, geographic and technological infrastructure.

The above review suggests that culture is a complex notion which is best assessed in terms of multiple dimensions. Markus and Soh (2002) suggested that in

REFERENCES

1. Barnes, S. & Huff, S.L., (2003) Rising Sun: iMode and the wireless Internet. *Communications of the ACM*, 46 (11), 78-84
2. Brown, I., R. Hoppe, P. Muger, P. Newman, and A. Stander (2004); 'The impact of national environment on the adoption of Internet Banking: comparing Singapore and South Africa'; *Journal of Global Information Management*; April-June v12, i2, pp1-26.
3. Burgess, A. (2004). *Cellular Phones, Public Fears, and a Culture of Precaution*. (1st Ed.). Cambridge.
4. Fife, E. & Pereira, F. (2003) The Diffusion of Mobile Data Applications. [Online paper]. Accessed on 27/03/04 from the World Wide web at: <http://www.marshall.usc.edu/ctm/publications/Research/diffusion%20of%20mobile%20data.pdf>
5. ForresterResearch <http://www.forrester.com/ER/Press/Release/0,1769,709,00.html>.
6. Geertz, C. (1973). "The interpretation of Cultures", Basic Books, New York, NY.
7. Heales, J. (2004); "National Culture, Infrastructure, and Costs: Factors that Lead to the Use of Wireless Technologies", *Proceedings of the Tenth Americas Conference on Information Systems*, New York, New York, August 2004.
8. Hofstede, G. (1997). *Cultures and Organizations: Software of the Mind*. (Revised Ed.). United Kingdom: McGraw-Hill International.
9. Hofvenschiold, E. (2003) Determining Cultural Issues in Attitude to and Use of Mobile Phones. *Proceedings of the 1st Annual GC-UPA Track*, Stuttgart, Germany, pg. 171-174.
10. International Business Center, (2003). [Online article] Accessed on 27/03/04 from the WWW
11. Javadian, M. (2001), "Organizational dimensions of global change: No limits to cooperation", *Administrative Science Quarterly*, (46-2), 354-356.
12. Jarvenpaa Sirkka L., Lang Karl Reiner, Takeda Yoko, and Virpi Kristiina Tuunainen; 'Mobile Commerce at Crossroads', *Communications of The ACM* December 2003/Vol. 46, No. 12.
13. Khalifa M., Cheng S.K.N.; 'Adoption of Mobile Commerce: Role of Exposure'; *Proceedings of the 35th Hawaii International Conference on System Sciences – 2002*.
14. Kluckhohn, F.R., and F.L. Strodtbeck. (1961), *Variations in Value Orientations*. Greenwood Press, Westport CT.
15. Lee, Y., Kim, J., Lee, I. and Kim, H. (2002). "A cross-cultural study on the value structure of mobile internet usage: comparison between Korea and Japan"; *Journal of Electronic Commerce Research*, Vol.3, No.4. 227-239.
16. Malone, T.W. (2001); "The Future of E-Business", *MIT Sloan Management Review*; Fall, Vol 43, Issue 1, pp104.
17. Markus, M.L., Soh, C. (2002) "Structural influences on global E-Commerce activity", *Journal of Global Information Management*, Jan-March, V10: 5(8).
18. Martin, R. (2003). *Germany: Expecting Growth*. [Online article]. Accessed on 27/03/04 on the World Wide Web.
19. Mizukoshi, Y., Okino, K. & Tardy, O. (2001) Lessons from Japan. [Online article]. *Telephony Online*. Accessed on 27/03/04 on the World Wide Web at: http://telephonyonline.com/ar/telecom_lessons_japan
20. Pousttchi, K. & Zenker, M. (2003). Current Mobile Payment Procedures on the German Market from the View of Customer Requirements. *Proceedings of the 14th International Workshop on Database and Expert Systems Applications*. Prague, Czech Republic pg.870-874
21. Samaras, G. (2002), 'Mobile Commerce: Vision and Challenges'; *Proceedings of the 2002 Symposium on Applications and Internet*.
22. Shannon, V. (2003) DoCoMo aiming to bring i-mode service to Britain. [Online article] Accessed on 27/03/04 on the World Wide Web at: <http://www.iht.com/articles/113527.html>
23. Straub, D. W. Loch, R. Aristo, E. Karahanna, and M. Strle, (2002); "Towards a Theory-Based Measurement of Culture", *Journal of Global Information Management*, Vol.10, No. 1: 13-23.
24. Sheth, D., W. Loch, R. Aristo, E. Karahanna, and M. Strite, "Towards a Theory-Based Measurement of Culture", *Journal of Global Information Management*, Vol. 10, 1:13-23, 2002.
25. Sweeney, J.C. and G.N. Soutar, "Consumer Perceived Value: The Development of a Multiple Item Scale", *Journal of Retailing*, Vol 77: 203-220.
26. Teng, J.T.C., Calhoun, K.J., Cheon, M., Raeburn, S. & Wong, W. (1999) Is the East really different from the West: A cross-cultural study on information technology and decision-making. *International Conference of Information Systems*, North Carolina, USA pg. 40-46
27. Tornatzky, L.G., and Fleischer, M. (1990) *The Processes of Technological Innovation*, Lexington, MA: Lexington Books.
28. Trompenaars, F. and C. Hampden-Turner, (1993); "Riding the Waves of culture", McGraw-Hill Inc., New York, NY.
29. Urbaczewski A., Wells J., Sarker S., Koivisto M., Amattikorkeakoulu M.; 'Exploring Cultural Differences as a Means for Understanding the Global Mobile Internet: A Theoretical Basis and Program of Research'; *Proceedings of the 35th Hawaii International Conference on System Sciences – 2002*.
30. Venkatesh Viswanath, Ramesh V., and Anne P. Massey; 'Understanding Usability in Mobile Commerce'; *Communications of The ACM* December 2003/Vol. 46, No. 12.
31. Warshauer, M. (2001). Singapore's Dilemma: Control vs. autonomy in IT-led development. *The Information society*, 17 (4), 305-311.
32. Xu, S., Zhu, K. and Kraemer, K.L. (2004), "Factors Affecting E-Business Diffusion in Organizations: Differences between Developed and Developing Countries"; *Proceedings of the Tenth Americas Conference on Information Systems*, New York, New York, August 2004.

Diffusion of Mobile Phones in Taiwan: An Evaluation of Influence Sources

Hersen Doong¹, Yicheng Ku²

¹ Department of MIS, National Chiayi University, Chiayi 600, Taiwan, China

² Department of Information Management, National Sun Yat-sen University, Kaohsiung 804, Taiwan, China
¹ hsdooong@mail.ncyu.edu.tw, ² yicheng.ku@msa.hinet.net

ABSTRACT

Mobile phones adoption and diffusion is a practical issue that has significant implications for e-commerce and m-commerce development. Examining the adoption of mobile phones from the well-developed theoretical foundation of innovation diffusion theory may clarify some significant factors that affect the adoption decision. Moreover, with the understanding of the influence sources of mobile phones adoption, the businesses can evaluate the number of mobile devices that can support the mobile commerce, plan the wireless capacity, formulate mobile commercial strategies, design more sophisticated applications, and launch more innovative wireless services.

In this research, we explore the sources of influence in the adoption of mobile phones in Taiwan, which mobile phone penetration ranks No.1 worldwide. We test adoption and diffusion pattern using three classical models: internal influence, external influence, and mixed influence models. Our findings suggest that mixed influence is the dominant influence factor in the diffusion of mobile phones. That is, the decision to become a new adopter of mobile phones is mainly determined by the positive influence of existing adopters and mass media channels.

Keywords: Innovation, Diffusion, Mobile Phone

1. INTRODUCITON

Explosive growth of the mobile phones over the past decade has greatly increased interest in its potential use of mobile commerce, especially as a vehicle for transactions. Plenty of mobile phones have provided consumers to perform various mobile commerce-related tasks without regarding to time and location. Therefore, tracking the adoption and diffusion pattern of mobile phones from the well-developed theoretical foundation of innovation diffusion theory is an increasingly important task, especially for mobile commerce planners. Examining the adoption of mobile phones from the well-developed theoretical foundation of innovation diffusion theory may clarify some significant factors that affect the adoption decision. With the understanding of the influence sources of mobile phones adoption and tracking the diffusion process, the businesses can evaluate the number of mobile devices that can support the mobile commerce's future use, plan the wireless capacity, formulate mobile commercial strategies, design more sophisticated applications, and launch more innovative wireless services.

Several behavioral models, such as Technology Acceptance Model, Theory of Reasoned Action, and Theory of Planned Behavior, for explaining and predicting the adoption and usage of technology have been proposed in the information systems literature. These studies primarily focused on surveys of individuals' perceptions but do not explore the sources of influence in the adoption of innovation within a social system. In this paper, we investigate the adoption process of mobile phones in Taiwan, which mobile phone penetration ranks No.1 worldwide, with

well-developed theoretical foundation of innovation diffusion theory. By means of the mathematical diffusion models, the process by which the innovation (mobile phones) is communicated through certain channels over time among the members of a social system can be analyzed. Three classical diffusion models [Loh et. al., 1992; Venkatrman, 1994; Hu et al., 1997], external influence model, internal influence model, and mixed influence model, are applied to depict the successive increase in the number of mobile phone adopters over time. The validation consequences will show us which factors influence the speed and specific course of mobile phone diffusion processes and may clarify some significant factors that affect the adoption decision. The research results will also allow us to have better understanding of the mobile phone growth, explore the sources of influence in the adoption of mobile phones, and illustrate possible implications on mobile commerce.

The rest of the paper is organized as follows: Firstly, we would like to briefly outline the current status of mobile communications in Taiwan. Then we will review the literature related to diffusion of innovation theory. Secondly, we would like to discuss the classical mathematical models of the diffusion of an innovation as developed by Mahajan and Peterson [1978, 1979, 1985]. At length, we will present the research results with longitudinal statistics and discuss their implications.

2. RESERARCH BACKGROUND

2.1. Mobile Phone Telephony

A mobile phone is a portable telephone which receives or

sends messages through base station network. Each base station overlaps other stations and has a range of 6-9 kilometers. Most people use the term “cellular phones” to mean all mobile phones using cellular technology, including GSM (Global System for Mobiles), CDMA (Code Division Multiple Access), DECT (Digital Enhanced Cordless Telecommunications), and traditional analog mobile technology. In a mobile phone network, users can roam with their mobile device and be handed off from one base station to the other. Mobile phones not only support voice calls, but send and receive data such as text, pictures, and short messages. Newer models also allow for accessing WAP (Wireless Application Protocol) services, have a built-in digital camera, and have some applications and entertainment programs integrated into them. At the present, mobile phones are changing to include the ability to access the World Wide Web and providing e-mail capability. They offer full Internet access using technologies such as GPRS (General Packet Radio Service).

2.2. The Status of Mobile Communications in Taiwan

The past decade has seen tremendous growth in the mobile sector in Taiwan. According to ITU (International Telecommunication Union) World Telecommunication Indicator Database, the number of mobile phone subscribers makes Taiwan to take the top seat in Asian mobile phone penetration ranking in 2000 and the 1st place in worldwide level in June of 2002. Relatively cheap rates, variety services, and aggressive operator subsidization programs have all contributed to Taiwan's mobile growth. With a population of about 23m, Taiwan is one of Asia's most dynamic markets. Figure 1 portrays the critical events and historical growth trend of mobile phone subscribers in Taiwan.

The high growth rate also means that service providers are facing a quickly maturing market. Despite aggressive marketing campaigns, operators are discovering that the pace of growth cannot be maintained. The market is quickly shifting from one focused on customer acquisition to one that is examining how to increase revenue with new service. The operators in Taiwan have begun to concentrate on improving earning ability base on its existing customer. At the present, the mobile commerce, mobile internet, short message service, WAP service, entertainment programs, and data/value-added services have been regarded as the major revenue sources of Taiwan's mobile phone operators. For example, data/value-added services usages last year grew 426% and 284% respectively. Moreover, personalization, such as interchangeable covers and user defined ring tones, has also helped in the uptake by the teen's market.

2.3. Diffusion of Innovation Theory and Diffusion Models

The diffusion of innovation theory provides an

explanation of how innovations diffuse through a social system. Diffusion is defined as the process by which an innovation is communicated through certain channels over a period of time among the members of a social system [Rogers, 1995]. Four key elements determine the characteristics of diffusion process of an innovation: innovation, time, social system, and communication channels. Rogers [1995] describe innovation as “any idea, object or practice that is perceived as new by the members of a social system... Time relates to the rate at which the innovation is diffused or the relative speed with which it is adopted by members of the social system... Social system consists of individuals, organizations, or agency ... are potential adopters of innovation... Communication channels are means by which information is transmitted to or within the social system”. As described above, communication channels have a significant effect on the diffusion pattern. Bass [1969] has categorized communication channels as either mass media messages or interpersonal communication channels. Mass media employ radio, television, newspapers, or other channels which enable a source of one or a few members within a social system to reach an audience of many. In contrast, interpersonal communication channels transmit an innovation to members within a social system via face to face communication between two or more individuals [Rogers, 1995]. Traditionally, the distribution of adopters over time is expected to be a bell-shaped curve. Adopters can be categorized into five categories by the different standard deviation from the mean: innovators, early adopters, early majority, late majority and laggards [Rogers, 1995].

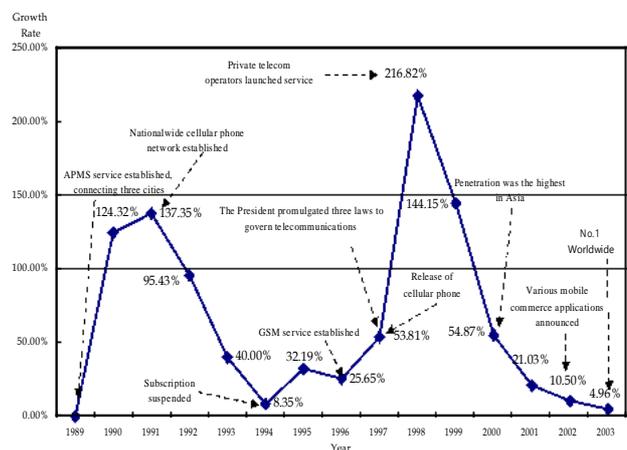


Figure 1. Historical Growth Rates

Since diffusion is the adopting process of the users among a social system, diffusion pattern, a line that is often a nonlinear curve, can describe the successive increase in the number of adopters or adopting units over time. Diffusion models are mathematical models that can be applied to depict the diffusion pattern, predict future distribution of an innovative technology, illustrate possible effect of a policy, or marketing of a new product. Mahajan and Peterson [1985] indicated three general

diffusion models: the external influence model, internal influence model and the mixed influence model. External influence model means the diffusion rate is affected by the mass media. The coefficient of diffusion is a constant, and the diffusion model was expressed as equation (1). Internal influence model hypothesize that the diffusion rate is affected by the interpersonal communication. Earlier adopters influenced the later adopters. The coefficient of diffusion is a function of the cumulative number of adopters, and the diffusion model was expressed as the equation (2). Mix influence model, equation (3), combined the effect of external and internal influence. All three general diffusion models are in the following:

$$N(t) = m(1 - \exp(-pt)) \quad \text{---- (1)}$$

$$N(t) = \frac{m}{1 + \frac{m - m_0}{m_0} \exp(-qmt)} \quad \text{---- (2)}$$

$$N(t) = \frac{m - \frac{p(m - m_0)}{p + qm_0} \exp(-(p + qm)t)}{1 + \frac{q(m - m_0)}{p + qm_0} \exp(-(p + qm)t)} \quad \text{---- (3)}$$

Where: $N(t)$ is the cumulative number of adopters,

$$N(t) = \int_{t_0}^t n(t) dt;$$

$N(t_0) = m_0$, the initial number of adopters in a social system;

$$\frac{dN(t)}{dt} = \text{diffusion rate at time } t;$$

m = total number of potential adopters in a social system.

In addition to three classical diffusion models, a rich body of studies adopt Exponential model to fit the diffusion pattern. Gurbaxani [1990] describes that an S-curve specification can be approximated by an exponential curve and characterized by a constant ratio of change when an innovation is in its early stages. Moreover, parameters of the diffusion models can be estimated in nonlinear estimation procedures [Srinivasan, 1986] or maximum likelihood estimation procedures [Schmittlein, 1982].

Diffusion models are built under some kind of patterns rather than random process. Therefore, it is necessary to verify the distribution of diffusion patterns after calculating the parameters. Mahajan et al. [1988] propose the White-Noise model, equation (4), to be the null hypothesis in order to test the validity of diffusion model. It means that diffusion model is useful when it can explain better than the White-Noise model on the innovation adoption process.

$$X(t) = x(t-1) + e(t) \quad \text{---- (4)}$$

Where: $x(t)$ is the number of adopters at time t ;

$$e(t) \text{ is assumed to be } N(0, s^2)$$

We adopt J-test [Davidson and Mackinnon, 1981] to test each influence model against White-Noise model, because White-Noise model is suitable for linear

problems and J-test was designed to test the specification of an econometric model in the presence of an alternative model that purport to explain the same phenomenon.

2.4. Studies of Innovation Diffusion

A rich body of studies has discussed innovation diffusion. For instance, Chaddha and Chitgopekar [1971] used a logistic formulation in estimating the demand for Bell System residence main telephones over the period 1966-1991. Wang and Kettinger [1995] applied a logistic model to survey the diffusion of number of subscribers and number of cell sites in the United States. The data they used covered the period from December 1984 to June 1994. The result shows that the logistic fit for number of cellular subscribers and the number of subscribers starts flattening out in year 2015 with 78,242,205 subscribers. The logistic model fit for number of cells as for the subscribers. The model also points out that a saturation point is reached at year 2020, with 33,248 cell sites.

Since the Internet is an example of a technology cluster innovation [Prescott, 1996], the prior researchers have done many investigations related to Internet diffusion. For example, Goodman et al. [1994] argued that most of the barriers to the wider distribution of networking tend to fall into three categories: government relative policies, technical impediments, and cultural factors. In addition to qualitative discussions, the quantitative methodology-mathematical diffusion models- are adopted to track the diffusion process of Internet hosts. For example, Gurbaxani [1990] found the pattern of diffusion in the case of BITNET is S-shaped and is consistent with a logistic curve, which describes a symmetrical growth process. Rai et al. [1998] adopted logistic, Gompertz and exponential models to forecast future development of Internet. They used nonlinear regression with data from August 1981 to January 1994 to estimate parameters for each of the three models. The result shows that the Exponential model provided a better fit than others in the initial growth stage of Internet. Moreover, Kauffman and Techatassanasoontorn [2004] used a modified Bass model and a coupled-hazard survival model to test the effects of country environmental factors, digital and analog wireless phone industry environmental factors, and technology policy factors on the speed of diffusion. They found that multiple standards and high prices slow down the digital wireless phone diffusion process from the Introduction state to the Partial Diffusion state. Competition in both analog and digital wireless phone industry also shapes the growth of digital wireless phone diffusion.

3. MODEL EXAMINATION

Mobile phone is an innovation of communication technology for people to communicate with each other. Therefore the number of mobile phone subscribers is influenced by both the number of existing subscribers

and the number of potential future subscribers. This section would explore the sources of influence in the adoption of mobile phones in Taiwan. We will test adoption and diffusion pattern of source influence using three classical models: internal influence, external influence, and mixed influence models.

3.1 Data

In order to estimate the parameters of diffusion model, we collected longitudinal statistics of mobile phone subscribers from 1989 to 2004 in Taiwan. The data source is ITU World Telecommunication Indicator Database. Table 1 shows the historical data of mobile phone subscribers in Taiwan. We use the data between 1989 and 2001 to estimate the parameters of three classical models and the data between 2002 and 2004 for checking sum of squares of the difference.

Table 1. The Historical Data of Mobile phone Subscribers in Taiwan

Year	Number of Subscribers	Year	Number of Subscribers
1989	37,000	1997	1,492,000
1990	83,000	1998	4,727,000
1991	197,000	1999	11,541,000
1992	385,000	2000	17,874,000
1993	539,000	2001	21,633,000
1994	584,000	2002	23,905,000
1995	772,000	2003	25,090,000
1996	970,000	2004	25,207,000

3.2. Testing Methods

In this section, we conduct two types of statistic tests to determine which model best fits the observations of mobile diffusion. Firstly, we estimate the model parameters of all three diffusion models with the nonlinear least squares regression method. Then we determine the true model among the three alternatives using the J-test. Finally we calculate the sum of the squares of the difference between predictive value and actual subscriber numbers.

NLIN, a nonlinear regression procedure of the SAS software package, is used to estimate the parameters in the three diffusion models. Table 2 summarizes the parameter estimation and the model fit for the three models. As Table 2 illustrates, the internal model has an adjusted R^2 of 0.9833 which is the proportion of variance explained by this model. The mixed model has an adjusted R^2 of 0.9800 and the external model just has an adjusted R^2 of 0.6687. That is, the adjusted R^2 of internal model is almost exactly the same as the adjusted R^2 of mixed model. Both of them are significantly higher than the R^2 -adjusted value of

external model. Thus, the overall predictive fit of external model is the worst.

Table 2. Parameter Estimation and Model Fit

Parameter	Internal Influence	Mixed Influence	External Influence
m	47,433	35,134	266,620
m_0	37	37	37
p		6.54E-21	0.0038
q	1.2433E-05	1.73683E-05	
R^2	0.9833	0.9800	0.6687

NOTE: 1.Unit: 1,000 subscribers

After the model parameters are estimated, the J-test is conducted to use each model in turn as the null hypothesis and the rest as the alternative hypotheses. Table 3 presents the results of comparisons of the alternative specifications with the White-Noise model with the J-test.

Table 3 J-test Model Comparisons: t Values with Probabilities

Null Model	Alternative Models		
	Internal Influence	Mixed Influence	External Influence
White Noise			
t statistic	8.3669	7.8965	0.7924
p-value	0.0000****	0.0000****	0.4449
**** p<0.001			

As table 3 shows, the J-test result rejects the null hypothesis that the adoption pattern of mobile phone is a random process. The alternative models can reject the White-Noise model at the significance level of $p < 0.001$ or better. Thus we come to a conclusion that the diffusion of mobile phones is not a random process. However, table 3 also indicates that the J-test result can't reject the null hypothesis while alternative model is external-influence. That is, the external model is not suitable for predicting the diffusion process of mobile phones. This also confirms the parameter estimation and model fit results described in Table 2. Accordingly, we don't analyze the external model any more in next section.

3.3. Best Model and Discussions

Figure 2 presents the curve of actual numbers of mobile phone subscribers and model forecasts. In order to examine the predictive validity of internal model and mixed model, we calculate the sum of squares of the differences between actual and predicted number of subscribers with the data from 2002 to 2004.

Table 4 summarizes the actual and predicted number of subscribers and sum of squares error for each model. As table 4 shows, the mixed model performs better than another one. Our research result clearly indicates that the mixed model best describes the diffusion process of mobile phones adoption. That is, the mixed influence is the dominant influence factor in the diffusion of mobile phones adoption. Hence, the decision to become a new adopter of mobile phones is mainly determined both by the positive influence of existing adopters and mass media messages. Therefore, mass media play a critical role in the earlier diffusion stage. These mass media publicize the advantage about using mobile phones widely in several years ago. However, the effect of mass media is not the only outstanding factor in the adoption process.

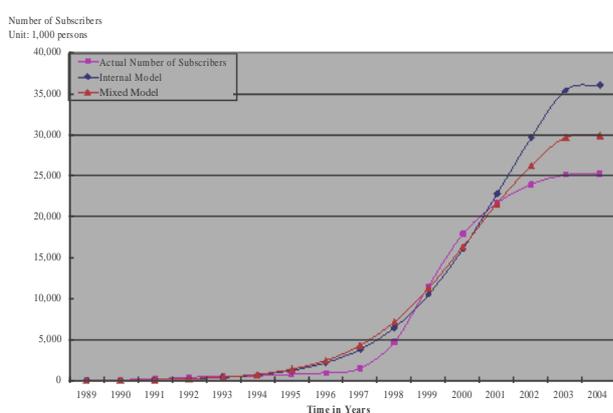


Figure 2. Actual Numbers of Mobile phone Subscribers V.S. Model Forecasts

Mahajan and Peterson [1985; p.18] noted that the internal influence perspective is “most appropriate when an innovation is ... socially visible, and not adopting it places social system members at a disadvantage”. With the rapid development of mobile technology, the mobile phone offers services independent of time and space. These benefits are perceived by users and are communicated to potential adopters. Even individuals will ask their friends, colleagues, or staffs to adopt mobile phones for communicating to each others more efficient. Besides, many individuals adopted mobile phones for the purpose of imitation. Factors which may motivate people to imitate buying mobile phones vary. For instance, factors considered in prior researches include perceptions that a user who adopts the mobile phone will be taken professional image, that a mobile phone makes me feel efficient, that it makes the user appear contemporary, dynamic and innovative.

Table 4. Model Forecasts vs. Actual Number of Subscribers

Year	Actual Subscribers	Internal Influence	Mixed Influence
1989	37	37	37
1990	83	67	68

1991	197	120	125
1992	385	216	230
1993	539	389	420
1994	584	696	766
1995	772	1,241	1,384
1996	970	2,192	2,467
1997	1,492	3,811	4,288
1998	4,727	6,457	7,159
1999	11,541	10,496	11,251
2000	17,874	16,072	16,317
2001	21,633	22,783	21,601
2002	23,905	29,647	26,213
2003	25,090	35,593	29,652
2004	25,207	36,024	29,883
Sum of squares of the difference		260,291,062	48,003,684

Furthermore, the pre-adopters may require friends of their social network to use mobile phones, especially the mobile phone has become a popular communication tool for most people. This is a strongly relationship for communication mechanisms where the value of the mobile phones increase with an increase in the number of users it connects. The mobile phone adoption is communicated among the members within the social system via interpersonal channels.

4. CONCLUSION

Mobile phone is one of the most important innovations over the last decade of the 20th century: with which communications become available anytime and any place. Accessibility, convenience, localization, and confidence make mobile services so unique. Mobility changes not only the life style of people but the enterprise’s business model. Who, then, pushes the growth of the mobile phones? Our study addresses this issue with the diffusion of innovation perspective. The research results show that mixed model, which assumes that the decision to become a new adopter is mainly determined by the positive influence of existing adopters and mass media channels, has better predictive validity. This leads us to conclude that it is the combined effects of external media, interpersonal pressure, and internal communications at the personal level that significantly influence the decision to adopt mobile phones. Moreover, network externalities might also offer a good explanation [Gurbaxani, 1990]. That is to say, the value of the mobile phones increases with an increase in the number of users it connects. Another explanation is that imitation within social system is a very popular phenomenon. The potential adopters perceive the value from early adopters and then decide to buy mobile

phones. While this study provides empirical result of diffusion process of mobile phones adoptions in Taiwan, future research can examine diffusion models with different country's data or apply the different model to have better understanding of mobile phones diffusion.

REFERENCES

- [1] Bass, F.M., "A New Product Growth Model for Consumer Durables," *Management Science*, Vol. 15 No. 5, pp. 215-217, 1969.
- [2] Chaddha, R.L., S.S. Chitgopekar, "A Generalization of the Logistic Curves and Long-range Forecasts (1966-1991) of Residence Telephones," *The Bell Journal of Economics and Management Science*, Vol. 2, No. 2, pp. 542-560, 1971.
- [3] Davidson, R., J.G. MacKinnon, "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica*, Vol. 49, No., pp. 181-793, 1981.
- [4] Goodman, S., L., S. Press, A. Ruth, A. Rutkowski, "The Global Diffusion of the Internet: Patterns and Problems," *Communications of the ACM*, Vol. 37, No. 8, pp. 27-31, 1994.
- [5] Gurbaxani, V., "Diffusion in Computing Networks: The Case of BITNET," *Communications of the ACM*, Vol. 33, No. 12, pp. 65-75, pp.1990
- [6] Hu, Q., C., Saunders, M. Gebelt, "Research Report: Diffusion of Information Systems Outsourcing: A Reevaluation of Influence Sources," *Information Systems Research*, Vol. 8, No. 3, pp. 288-301, 1997.
- [7] Kauffman, R. J., A. A. Techatassanasoontorn, "Does One Standard Promote Faster Growth? An Econometric Analysis of the International Diffusion of Wireless Technology," *Proceedings of the 37th Hawaii International Conference on System Science*, 2004.
- [8] Loh, L. N. Venkatraman, "Diffusion of Information Technology Outsourcing: Influence Sources and the Kodak Effect," *Information Systems Research*, Vol. 3, No. 4, pp. 334-358, 1992.
- [9] Mahajan, V. R., A. Perterson, "Innovation Diffusion in a Dynamic Potential Adopter Population," *Management Science*, Vol. 24, pp.1589-1597, 1978.
- [10] Mahajan, V., R. A. Perterson, "Integrating Time and Space in Technological Substitution Models," *Technological Forecasting and Social Change*, Vol. 14, pp. 127-146, 1979.
- [11] Mahajan, V., R. A. Peterson, *Models For Innovation Diffusion*, Beverly Hill, CA: Sage Publication Inc., 1985.
- [12] Mahajan, V., S. Sharma, R.A. Bettis, "The Adoption of the M-Form Organizational Structure: A Test of Imitation Hypothesis," *Management Science*, Vol. 34, No. 10, pp. 1188-1201, 1988.
- [13] Prescott, M.B., C. V. Slyke, "The Internet as an Innovation," *Proceeding of the Association of Information Systems Conference*, 1996.
- [14] Rai, A., T. Ravichandran, S. Samaddar, "How to Anticipated the Internet's Global Diffusion," *Communications of the ACM*, Vol. 41, No. 10, pp. 65-75, 1988.
- [15] Rogers, E.M., *Diffusion of Innovations*, 4th ed., New York, 1995.
- [16] Rohlfs, J., "A Theory of Interdependent Demand for a Communications Services," *The Bell journal of Economics and Management Science*, Vol. 5, No. 1, pp. 16-37, 1974
- [17] Schmittlein, D., V. Mahajan, "Maximum Likelihood Estimation for an Innovation Diffusion Model of New Product Acceptance," *Marketing Science*, Vol. 1, No. 1, pp. 57-78, 1982.
- [18] Srinivasan, V., C. H. Mason, "Nonlinear Least Squares Estimation of New Product Diffusion Models," *Marketing Science*, Vol. 5, No. 2, pp. 169-178, 1986.
- [19] Venkatraman, N., L. Loh, J. Koh, "The Adoption of Corporate Governance Mechanisms: A Test of Competing Diffusion Models," *Management Science*, Vol., 40, pp. 497-506, 1994.
- [20] Wang, M., W.J. Kettinger, "Projecting the Growth of Cellular Communication," *Communication of the ACM*, Vol. 38, No. 10, pp. 119-122, 1995.

Explaining Intent of Adoption of Wireless Internet Services by Business Executives: The Case of Paris Airports' Hotspots

Madeleine Besson, Olivier Epinette, Olivier Segard, Pierre Vialle

Institut National des Télécommunications, Evry, 91011, France
{Madeleine Besson, Olivier Epinette, Olivier Segard, Pierre Vialle}@int-evry.fr

ABSTRACT

According to most industry experts and researchers, wireless Internet access is seen as a promising market for mobile users in locations such as airports and other public venues [1]. According to the Gartner Group, 6,000 wireless Internet hotspots were installed in the world in 2002, 57% in Northern America and 14% in Europe. 15,000 hotspots should have been implemented in Europe by 2000, used by 11 million regular users and 12 million occasional users. Airports, because of the high density of travellers and the relatively long waiting times involved, are considered as key hotspot locations. Yet, consumption of such services is still lagging behind suppliers' expectations. Why, and whether, consumers will use wireless Internet still remains unclear. This paper aims to contribute to answer both questions by defining a framework explaining the intent of use by consumers. To this end, we carried out market research, focused on executives attending public venues where wireless Internet services can be proposed. In this paper we first define the specific features of the wireless Internet access for airports. We then propose a framework explaining the intent to use wireless Internet services.

Keywords: e-business, m-business, product adoption, telecommunications, mobile Internet, wireless Internet

INTRODUCTION

Understanding and anticipating potential consumer behaviour is critical for suppliers of new high technology services. However, this is a particularly difficult exercise because of factors such as the complexity of the products involved, the presence of interdependencies, the low or non-existing consumer experience, or the high incertitude of the impact of innovations.

More specifically, consumer behaviour concerning wireless Internet services in public venues is hard to understand for the following reasons. First, these services can be used by a large variety of travelling professionals, even if senior executives currently constitute the main target of hotspots. Second, adoption and usage are dependent both on the decision of the individual users, and of the company employing them. Third, from the individual user's perspective, there is no clear border between professional and more personal expectations. Family, entertainment and business topics must be equally considered. Fourth, a multitude of companies can coexist on a hotspot, as in the case of an airport. So, the spectrum of localised mobile Internet services, which can be proposed, is broad. Fifth, the adoption and consumption of such services are subject to network externalities: consumption in one hotspot is dependent on consumption in other geographical areas. Finally, beyond the services to be provided, the physical characteristics of the hotspot could have an influence on the consumption experience.

To explore this complex consumption situation, we conducted a survey in co-operation with the state-owned organisation which manages the main airports near Paris.

From this survey, we present, in the second part of this paper, a framework aiming at explaining the intent to use wireless Internet services. We begin with the context of the survey.

1. THE CONTEXT: WIRELESS INTERNET HOTSPOTS IN AIRPORTS

The research¹ was conducted in the main airports near Paris, in order to analyse the intent to use Wireless Fidelity (Wi-Fi) services in these hotspots, the provision of which began in the year 2003.

An airport covers different activities and geographic zones. Several activities can be carried out on an airport platform, such as transport, control, cleaning, maintenance, or passenger care. Geographically, these activities broadly relate to three zones on the airport area: tarmac (services concerning aeroplanes), air terminals (services for passengers on site) and airport periphery (activities benefiting from close location to the airport, such as hotels or offices). Finally, three categories of people can be found on an airport site: employees of businesses and administrations located on the airport platform, passengers, and employees from outside firms on short duty at the airport (for example for repair activities). In this study, we focus on business travellers that have been identified as the main market target by ADP, the organisation managing the various airports around Paris.

¹ This research is drawn from a broader research project called ERACE, including also an inter-organisational component [3].

The study relates to a hotspot service in the main airports of Paris, using a Wireless Local Area Network (WLAN) technology. WLAN is a short-range, broadband data communication system using radio waves [2]. Various standards coexist, but the 802.11b standard, defined by the Institute of Electrical and Electronics Engineers (IEEE) tends to be dominant. It allows a theoretical flow of 11 Mb/s in the frequency band of the 2.4 GHz, on a distance of up to several hundred meters. The IEEE 802.11b standard is more popular under the name of Wi-Fi. The deployment of a Wi-Fi network is relatively easy and it offers a much higher level of flexibility than fixed networks. The number of accesses can be increased without needing to re-dimension the network, as would be the case with fixed networks. This system allows some limited mobility inside one cell, but contrary to cellular networks such as GSM networks, does not allow roaming between cells without service interruption. Such networks can be accessed by different types of terminals equipped with Wi-Fi chips, such as mobile phones, Personal Digital Assistant (PDA), or laptop computers.

There are three possible uses of Wi-Fi. The first is private use, either for a home network, or for a company Local Area Network (LAN). The second is community networks, usually in rural areas. The third is hotspots in public venues such as coffee shops, hotels, railway stations or airports.

A Wi-Fi network can give access to the following services:

- Internet access: including surfing, messaging and Virtual Private network services allowing access to Intranet,
- Localised information services: reception of "push" information, such as infrastructure usage information and schedules (transport hotspots), or commercial information (for example from nearby shops),
- Multimedia services, such as movies or music. As the available bandwidth is shared between current users, a large number of users can result in unacceptable download time.

As users of a particular hotspot may be customers of another hotspot, or of a mobile operator, the establishment of roaming agreements is necessary. Basically, a roaming agreement ensures that a customer of network A is able to use the service provided by network B, while being billed by network A, with some mechanism of revenue sharing between the two networks. Therefore, users can pay for service usage in three different ways. They can be billed by the hotspot manager, either directly through payment by credit card,

or indirectly with prepaid cards that can be bought in the airport's newspaper shops. They can also be billed on their mobile phone account, as most mobile operators offer subscription schemes bundling General Packet Radio Service (GPRS) cellular services with Wi-Fi services.

2. THE FRAMEWORK: FACTORS EXPLAINING INTENT TO USE

To explore this complex subject, we made a qualitative study, and not a quantitative research, which would be limited to quantifying traditional uses, and might not take into account all the richness of the consumption experience in a hotspot, and of the psychosocial context. The qualitative study is based on twenty-four interviews with medium and senior executives. From our preliminary results, we were able to conceive a simple framework aiming at explaining the intent of adoption and use of wireless Internet services. This framework specifically concerns an airport environment, but can easily be adapted to other types of public hotspots.

The sampling frame was made up of several lists of business schools' alumni. People were selected in order to ensure the strongest variety of the points of view on our study subject. The selection was operated starting from two variables that could discriminate the answers: the function in the company and the degree of mobility. After having contacted further these two hundred people, we were able to carry out twenty-seven interviews. Three could not be used because of bad audio recording quality.

To collect information, individual in-depth interviews were carried out. They are appropriate for collecting information concerning perception and behaviour. The interviewed person expressed himself freely, but within the framework of the topics suggested by the guideline. The former is composed of three parts. The first part makes it possible for the person interviewed to evacuate the stereotypes concerning the Internet and to centre on the current uses. The second part centres the discussion on travel and the situations of waiting able to occur on this occasion. The last part proposes to imagine the mobile Internet services able to be offered.

Four categories of factors emerge from our qualitative analysis, namely individual characteristics and experience of business travellers, the policy of their organisation concerning Information and Communications Technology (I&CT), situational factors relative to the hotspot, and service availability inside and outside the hotspot (table 1).

Table 1. Factors explaining intent to use Wireless Internet in an airport

<i>Factors</i>	<i>Explanations</i>
<i>Individual characteristics and experience</i>	
Position	Managers with higher positions are more likely to be equipped, but less likely to use
Function	Mobile communications are more crucial for some functions (e.g. sales, technical support)
Age	Older people are less likely to use
Sex	Men more likely to use than women
Personality	Desire to keep in touch, need for anticipation and reassurance
Experience	Experience in using the Internet and mobile devices (learning effect, familiarity), related to personal interest, technical functions, sector of industry
User status	Effect of positive/negative experience on general attitude towards computing and Wi-Fi Wi-Fi user in other contexts
<i>Organisation policy</i>	
Attitude towards I&CT	Strategic resource or cost to minimise
Equipment and access policy	Influence equipment
Service policy	Policy to spur or restrict usage
Security policy	May lead to restrict availability of remote access to IS
<i>Situational factors</i>	
Perceived available time	The more perceived available time, the more usage
Travel characteristics	The type of travel, in terms of duration, destination, transit, determines the available time. A low level of organisation induces need for information and transaction
Physical environment	Privacy, comfort, noise level, availability of sockets
<i>Service availability</i>	
At hotspot	Attractiveness and width of services available influence usage
At other hotspots	The more hotspots, the more adoption and incentive to use (network externalities)
At company premises	Availability of company WLAN induces adoption and propensity to use
Roaming availability	Reduce transaction costs, induce spontaneity of usage

2.1 Individual characteristics and experience

Various factors seem to have an influence on the propensity to use wireless Internet in a hotspot: hierarchical position, function, age, sex, personality, experience and user status.

The hierarchical position plays an ambiguous role. On the one hand people with a higher position are more likely to be equipped with terminals by their organisation, and particularly with PDAs. On the other hand, their comments suggest that travel time is often perceived as an opportunity to read, think and relax, rather than dealing with a computing device.

The function in the organisation also has an influence on intended usage. First, for some functions, such as sales or technical support, communicating in situations of mobility can be critical. People carrying on such functions are therefore more likely to be equipped, to be experienced in communicating while being mobile, and to have urgent needs in a hotspot. In general, executives working in technical functions or in the I&CT sector are more experienced in using computing devices and therefore more likely to use Wi-Fi services.

Age can have a low influence on usage, as propensity to use seems to decrease with age.

Men exhibit a higher propensity to use wireless Internet than women. One explanation could be that laptop computers and their connection do not appear practical.

PDAs are considered as an interesting alternative, but are still not perceived as user friendly enough by women.

Some personality features seem to influence usage, and particularly one characteristic that we call "need for reassurance". Even when the trip is organised in advance, some respondents expressed the need to connect to the Internet in order to check information or to be able to deal with unexpected events, concerning, for example, a hotel or car reservation. The role of "orientation towards communicating with others" can be ambiguous. On the one hand, such people can be expected to communicate electronically, on the one hand they may prefer to communicate directly, for example, with their neighbour. The latter case seems to be more likely.

Experience plays a role through two distinct but related aspects. The first concerns the accumulated experience, and hence familiarity and skills with the use of computing and telecommunications products and services. The more experienced the user, the higher the propensity to use other products and services such as wireless Internet. This is apparent, as explained before, through the type of function (e.g. technical), or the sector of industry. The second concerns the positive or negative attitude generated by former experience. Interestingly enough, some respondents having tested Wi-Fi in another context (WLAN) had a negative attitude due to the problems they encountered.

User status: being already a Wi-Fi user in other hotspots naturally leads to an increase in the propensity to use it in an airport.

2.2 Organisation policy concerning Information & Communications Technology

For business travellers the policy of their company or organisation concerning I&CT can have a decisive impact on the availability of equipment and of usage. In [5] and [6], we identified that the adoption behaviour for new I&CT products and services was strongly dependent on the way these investments and expenses were considered, and that Information Technology (IT) managers found it difficult to assess the return on investment for this type of project. The attitude towards I&CT investment and expenses broadly oscillates between two extreme positions. One possible attitude is to consider I&CT as a strategic investment (at least for a given number of projects), whose benefits are localised outside the IT function, such as improving internal and external co-ordination, reducing time to market, or increasing customer satisfaction. Another attitude is to treat these resources as a cost to be minimised, according to the available budget. In this case, investments will be considered if benefits are localised inside the IT function, such as a lower cost per bit, or lower maintenance cost.

The attitude towards I&CT will impact on the equipment and access policy (deciding who will be equipped with a terminal equipment and allowed to access a service), as well as the service policy, such as the range of services available.

Another critical point is the security policy of the IT department. A tight security policy may lead IT departments to restrict remote access to their Information System, in order to avoid malevolent intrusions (for example by hackers). IT departments may be even more cautious in the case of Wi-Fi access, as intrusion is easier to achieve for radio links than for fixed lines. An intruder could pretend to be another user for the server, or alternatively appear as the server to a user.

2.3 Situational factors

The main influencing factors are the perceived available time, the travel characteristics, and the characteristics of the physical environment. The perceived available time is relative in two aspects: the user personality (e.g. tolerance of inactivity), and the competing entertainment or other activities at hand.

Travel characteristics, in terms of duration, destination and transit time, have an influence on both time available and need for communication and information. For example, a passenger transiting between two intercontinental flights would have a high need to keep informed about his family or his business, and would also have enough time available for it. A low level of

pre-organisation of travel should also induce stronger needs for communicating, for example in order to confirm a reservation, or to get additional tourist information.

Consumption of Wi-Fi services should also depend on the physical environment. This refers to ambience factors such as benefiting from a comfortable enough environment, in terms of furniture, silence, and isolation from other people. It also refers to more critical issues such as the availability of sockets for terminal equipment to be plugged in to. These factors appear to be critical and also difficult to provide, as the organisation of space in an airport is strongly determined by considerations of security, passenger management, and scarcity of available space.

2.4 Service availability

The expected consumption will be dependent on the availability of attractive services both inside and outside the hotspot considered. Service expectations are rather standard and mainly concern e-mail and Internet access. For PDA users, the provision of information in a suitable format is required. In terms of content, respondents seem to focus more on travel related information. "Push" information, for example to keep informed about accurate boarding and departure time, also raised interest. We did not identify much attraction towards multimedia services, such as downloading movies².

Expected consumption is also related to service availability outside the hotspot considered, through the number of other hotspots, and/or the existence of WLANs in the organisation employing travellers. This is a typical example of network externalities where the utility of a service is dependent on the availability of complementary products and services [4]. In fact it is unlikely that potential users would adopt Wi-Fi specifically in the case of the Paris airports. We would expect a two-stage decision process: first a decision to adopt Wireless Internet at a general level, or in a company WLAN, and then to use it in the airport under a given number of conditions. Multi-usage means the existence of roaming agreements between the various hotspot owners and mobile operators, including the possibility to be billed by one's usual provider.

CONCLUSION

We proposed a framework explaining the use of wireless Internet services by individual characteristics, company policy, usage situation, service availability and travel characteristics.

² However, this may result from a possible bias in the interviews, as we felt that respondents were rather reluctant to mention entertainment while talking about professional travel.

A first key result of our research is the importance of the consumption situation, such as the availability of electrical connectors, a comfortable environment, or the perception of the serviceable time. The provision of localised services and of a related physical environment seems to be able to trigger the consumption of wireless Internet services on hotspots.

A second key result is that focusing only on the individual's needs and expectations, a traditional characteristic of consumer marketing, is not sufficient in order to understand the product's potential for adoption. The individual adoption behaviour will be obviously influenced by the company's computer policy, such as the terminal equipment policy, the security policy or the services provided by the organisation.

A third key result consists of the confirmation of the importance of network effects. The consumption of services in a specific hotspot is dependent on the consumption in other places, related to geographical range of availability of similar services, as well as the provision of adequate billing solutions.

The next step of our work is to use the framework to develop a set of hypotheses to be tested with a quantitative study.

REFERENCES

- [1] Besson, M., O. EpINETTE, O. Segard, "Will the Wireless Internet Market "Take off"? The Case of the French Business Passengers", *the 2004 Air Transport Research Society World Conference (ATRSWC 2004)*, Istanbul, Jul. 1-3, 2004.
- [2] Dineen, R., S. Anderton, *Public wireless LAN: the business opportunity*, Ovum, 2002.
- [3] EpINETTE, O., O. Segard, P. Vialle, "The structuring of a value chain in order to deliver new system-based services: the case of wireless Internet for Paris airports", *proceedings of The International Conference on Service Systems and Service Management (ICSSSM04)*, Vol.1, pp. 872-876, Beijing, Jul. 19-21, 2004.
- [4] Katz, M., C. Shapiro, "Network Externalities, Competition and Compatibility", *American Economic Review*, Vol. 75, pp424-440, 1985.
- [5] Vialle, P., "The emergence of m-business: an analysis of the case of the French market", *proceedings of 14th European ITS (International Telecommunications Society) Conference*, Helsinki, Aug. 23-24, 2003.
- [6] Vialle, P., O. EpINETTE, "The attitudes and perceptions of ICT managers of large firms towards M-business Internet adoption: beyond the magic...", *proceedings of Conference on Telecommunications and Information Markets (COTIM)*, Karlsruhe, Jul. 18-20, 2001.

Extracting Requirements for Ubiquitous Computing Technology-based IS Using Factor Analysis

Natalia Em, Euiho Suh, Keedong Yoo

Dept. of I.E., Pohang University of Science and Technology, 790-784, Republic of Korea
{natalia, ehsuh, kdy}@postech.ac.kr

ABSTRACT

The paper describes a methodology for extracting the requirements for ubiquitous computing technology-based information system (IS) using factor analysis. The early idea of ubiquitous computing was articulated by Mark Weiser in 1988 and since then has been constantly developed. However, so far there was no any or few research on application of ubiquitous computing to IS. Prior to the development of a ubiquitous computing-based IS model necessary requirements are to be identified. To derive the requirements we refer to conventional research on ubiquitous computing, ubiquitous service and ubiquitous environment, and hence, to their characteristics. Having specified those requirements using factor analysis technique it will be easier to develop a consistent ubiquitous computing-based IS.

Key words: ubiquitous computing, information system (IS), factor analysis.

1. INTRODUCTION

In this paper we attempt to develop a methodology in order to identify the requirements for ubiquitous computing-based IS with the help of data analysis. Numerous projects and research were done in the field of ubiquitous computing, service and environment since the first attempts to identify ubiquitous computing concept have been undertaken by Mark Weiser. Conventional research can be categorized into four main areas: U-Device, U-Environment, U-Media and U-People [14]. With ubiquitous computing advent research was mainly focused on creating a proper hardware such as active badges, beacons, smart tags (U-Device). Meanwhile, other explorations were undertaken in developing ubiquitous computing technology-enabled environments, so called "smart environments" (U-Environment). Yet, U-Media area has not progressed as much because of "users' often conservative attitude toward new media, as well as the technical complexity in fitting the new types of media to users"[15]. Despite some technical complexity, researches concentrated on the development of new concepts of interface patterns for users with disabilities are conducted in U-People field.

However, only few works have been carried out towards the connection of the notions of ubiquitous computing and IS so far. One can try to develop a model of IS based on ubiquitous computing technology. Still, before one does this, the requirements for such a model are to be defined. Otherwise, ignoring the requirements can lead to development of an inconsistent model.

In order to derive the requirements specific for the development of ubiquitous computing-based IS related

literature, namely, conventional literature on ubiquitous computing must be examined. Having reviewed the literature we then outline certain characteristics and

features that are intrinsic to ubiquitous computing. Out of the multiple characteristics it becomes possible to determine particular keywords that can serve as a platform for formulating necessary requirements for ubiquitous computing-based IS.

For obtaining the requirements a formal tool should be used. Factor analysis is one of them since it is a statistical tool that is determined to reduce the number of variables as well as to detect structure in the relationships between variables.

The remainder of the paper is organized as follows. Section 2 reveals related works. Section 3 is dedicated to the methodology. Finally, section 4 concludes the paper with summary and future work.

2. RELATED RESEARCH

2.1 Ubiquitous Computing

The concept of ubiquitous computing was first proposed by Mark Weiser in early 1990s. In his early works he describes ubiquitous computing or "Ubicomp" as a way to augment computer usage by "making many computers available throughout the physical environment, but making them effectively invisible to the user" [12, 13]. The technology requirements comprise three aspects: cheap, low-powered computers with convenient displays, a network that unites them as well as software systems implementing ubiquitous applications [14].

Hence, multiple issues have arisen with the advent of ubiquitous computing, particularly related with the development of proper hardware and devices, software and applications, environment, etc. Beader et al. [2] in their research investigate the usage of mobile location aware systems equipped with Badge based location and environment sensors where they tried to overcome the

deficiencies connected with environment awareness. Xerox PARC carried out explorations of the capabilities of mobile computers in an office setting which have resulted in the development of PARCTAB system prototype [11]. Narumi et al. [9] pose the technology requirements for ubiquitous environment and describe Fujitsu's real-time communicator ubiquitous middleware. Meanwhile, the FUSE (Future Ubiquitous Service Environment) group's focus is software architectures for ubiquitous service environments designing as well as privacy and context sensitivity [8]. Epp [5] stresses privacy and authenticated identity issues within ubiquitous environment. Burrell et al. [3] discuss the design and development of Semaphore, a contextually aware application for use in wireless networked environments.

2.2 Ubiquitous IS

As it was mentioned above, only few explorations have been done in this direction. However, we give a brief description of more or less corresponding works. "Docodemo-Display" is one of them and is a recently developed ubiquitous IS. It is a display-centric communication system, which is portable, interactive, and collaborative [10]. The goal of "Docodemo-Display" System is to increase the value of the place with helping to form local communities at that place. Another research is by Balke et al. [1]. In their work "A Situation-aware Mobile Traffic Information System" they present a prototype of traffic IS offering advanced personalized route planning, including added service like traffic jam alerting by means of SMS.

2.3 Factor analysis and Its Applications

Factor analysis is a branch of applied mathematics [6] and is generally referred to as a statistical tool for reduction of the number of variables and structure detection in the relationships between variables. First applied in psychology this method quickly became popular in other fields like economics, medicine, sociology, etc. as they often need to outline dominating factors out of huge volumes of variables they have to operate with.

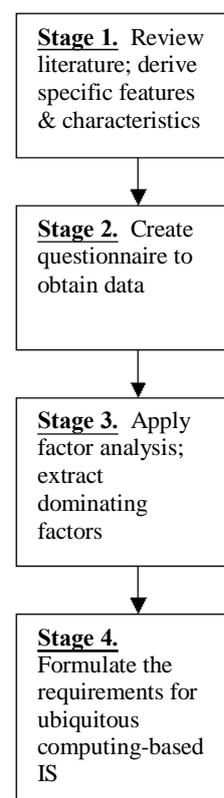
One of the conventional works using factor analysis describes the application of the latter to tests on Cheshire cheese [7]. The study reveals the relationships between a number of rheological experiments and essential subjective judgments on cheese. Another research applies factor analysis to psychological diagnostics [4] and particularly to the development of psychological tests and models such as the Cattell 16 Personal Factors model.

Numerous studies in various fields using factor analysis can be found. However, no studies seem to be carried out towards extracting the requirements for ubiquitous computing technology-based IS using factor analysis.

3. METHODOLOGY

Owing to well-developed concept we find factor analysis applicable to extracting essential requirements for ubiquitous computing-based IS. However, due to impossibility to collect data necessary for conducting empirical tests in a short period of time we describe only a methodology that comprises several stages (Figure 1). The first stage considers reviewing conventional literature in ubiquitous computing, services or environment fields and deriving specific features and characteristics. Next, on stage two we create a questionnaire for the experts in the related fields. The questionnaire will include the list of the derived features and characteristics that must be evaluated by the experts in terms of their significance. Having done this, we will get a data sample. The data will perform experts' estimations of the characteristics in scores. Having obtained the essential data we then apply factor analysis on stage three and extract dominating factors. Eventually, on stage four, relying on the results of factor analysis we formulate the requirements for ubiquitous computing-based IS.

Figure 1. Methodology of extracting the requirements to ubiquitous computing-based IS.



4. SUMMARY AND FUTURE WORK

The fact that very few works can be found on application of ubiquitous computing to IS and, therefore, the requirements for such an IS are not clearly articulated, makes this paper unique for it aims to

extract and classify those requirements. In this paper we delineate a general methodology of extracting the requirements using such a common statistical tool as factor analysis.

The future work within the research will lie in applying the described methodology to practice. Probably the most time-consuming stage will be collecting necessary data as it requires direct participation of respondents. However, having done this, it will be possible to acquire significant results which, in turn, will make a valuable contribution in ubiquitous computing-base IS development.

REFERENCES

- [1] Balke, W.-T., Kiebling, W., Unbehend, C., "A Situation-aware Mobile Traffic Information System", *proceedings of the 36th Hawaii International Conference on System Sciences*, 2002
- [2] Beadle, Peter H. W., Maguire Jr., G. Q., Smith, M. T., "Location Base Personal Mobile Computing and Communication", http://www.it.kth.se/~maguire/LocationAware/ieee_lan_98/, 1998
- [3] Burrell, J., Treadwell, P., Gay, Geri K., "Designing for Context: Usability in a Ubiquitous Environment", http://www.hci.cornell.edu/LabArticles/DesigningContext_Burrell.pdf
- [4] Chaplina, G., "Application of Factor Analysis in Psychological Diagnostics", <http://www.monax.ru/eleme.cgi?1889>, 2001 (in Russian)
- [5] Epp, E. C., "Relationship Management: Secure Collaboration in a Ubiquitous Environment", *Pervasive Computing Mobile and Ubiquitous Systems*, Apr-June 2003
- [6] Harman, H. H., *Modern Factor Analysis*, Chicago: The University of Chicago Press, 1976
- [7] Harper, R., Baron, M., "The Application of Factor Analysis to Tests on Cheese", *British Journal of Applied Physics*, Vol., 2, pp 35-41, 1951
- [8] Jansson, C. G., Lonnqvist, P., "Ubiquitous Service Environments", *ERCIM News*, No. 47, October 2001
- [9] Narumi, K., Yuhara, M., Matsumoto, T., "Technology Requirements for Ubiquitous Environment and Fujitsu's Realtime Communicator Solution", *Fujitsu Sci. Tech. J.*, 40, 1, pp 29-34, June 2004
- [10] Sekiguchi, T., Iguchi, S., Ishii, M., Maeoka, J., Ito, A., Furuhashi, T., "A Display Centric Ubiquitous Information System-"Docodemo-Display", *Symposium on Applications and Internet-WorkShops (SAINT 2004 Workshops)*, <http://csdl.computer.org/comp/proceedings/saint-w/2004/2050/00/20500709abs.htm>, January 2004
- [11] Schilit, B., Want, R., "The Xerox PARCTAB", <http://sandbox.parc.xerox.com/parctab/parctab.html>, 1995
- [12] Weiser, M., "Hot Topics: Ubiquitous Computing", *IEEE Computer*, October 1993
- [13] Weiser, M., "Some Computer Science Issues in Ubiquitous Computing", *Communications of the ACM*, July 1993
- [14] Weiser, M., "The Computer for 21st Century", <http://www.ubiq.com/hypertext/weiser/SciAmDraft3.html>
- [15] Yoo, K., Kwon, O., Suh, E., "UbiDSS: A Proactive Intelligent Decision Support System as an Expert System Deploying Ubiquitous Computing Technologies", 2004 (to be appeared)

Extracting Requirements for Ubiquitous Computing Technology-based IS Using Factor Analysis

Natalia Em, Euiho Suh, Keedong Yoo

Dept. of I.E., Pohang University of Science and Technology, 790-784, Republic of Korea
{natalia, ehsuh, kdy}@postech.ac.kr

ABSTRACT

The paper describes a methodology for extracting the requirements for ubiquitous computing technology-based information system (IS) using factor analysis. The early idea of ubiquitous computing was articulated by Mark Weiser in 1988 and since then has been constantly developed. However, so far there was no any or few research on application of ubiquitous computing to IS. Prior to the development of a ubiquitous computing-based IS model necessary requirements are to be identified. To derive the requirements we refer to conventional research on ubiquitous computing, ubiquitous service and ubiquitous environment, and hence, to their characteristics. Having specified those requirements using factor analysis technique it will be easier to develop a consistent ubiquitous computing-based IS.

Key words: ubiquitous computing, information system (IS), factor analysis.

1. INTRODUCTION

In this paper we attempt to develop a methodology in order to identify the requirements for ubiquitous computing-based IS with the help of data analysis. Numerous projects and research were done in the field of ubiquitous computing, service and environment since the first attempts to identify ubiquitous computing concept have been undertaken by Mark Weiser. Conventional research can be categorized into four main areas: U-Device, U-Environment, U-Media and U-People [14]. With ubiquitous computing advent research was mainly focused on creating a proper hardware such as active badges, beacons, smart tags (U-Device). Meanwhile, other explorations were undertaken in developing ubiquitous computing technology-enabled environments, so called "smart environments" (U-Environment). Yet, U-Media area has not progressed as much because of "users' often conservative attitude toward new media, as well as the technical complexity in fitting the new types of media to users"[15]. Despite some technical complexity, researches concentrated on the development of new concepts of interface patterns for users with disabilities are conducted in U-People field.

However, only few works have been carried out towards the connection of the notions of ubiquitous computing and IS so far. One can try to develop a model of IS based on ubiquitous computing technology. Still, before one does this, the requirements for such a model are to be defined. Otherwise, ignoring the requirements can lead to development of an inconsistent model.

In order to derive the requirements specific for the development of ubiquitous computing-based IS related

literature, namely, conventional literature on ubiquitous computing must be examined. Having reviewed the literature we then outline certain characteristics and

features that are intrinsic to ubiquitous computing. Out of the multiple characteristics it becomes possible to determine particular keywords that can serve as a platform for formulating necessary requirements for ubiquitous computing-based IS.

For obtaining the requirements a formal tool should be used. Factor analysis is one of them since it is a statistical tool that is determined to reduce the number of variables as well as to detect structure in the relationships between variables.

The remainder of the paper is organized as follows. Section 2 reveals related works. Section 3 is dedicated to the methodology. Finally, section 4 concludes the paper with summary and future work.

2. RELATED RESEARCH

2.1 Ubiquitous Computing

The concept of ubiquitous computing was first proposed by Mark Weiser in early 1990s. In his early works he describes ubiquitous computing or "Ubicomp" as a way to augment computer usage by "making many computers available throughout the physical environment, but making them effectively invisible to the user" [12, 13]. The technology requirements comprise three aspects: cheap, low-powered computers with convenient displays, a network that unites them as well as software systems implementing ubiquitous applications [14].

Hence, multiple issues have arisen with the advent of ubiquitous computing, particularly related with the development of proper hardware and devices, software and applications, environment, etc. Beader et al. [2] in their research investigate the usage of mobile location aware systems equipped with Badge based location and environment sensors where they tried to overcome the

deficiencies connected with environment awareness. Xerox PARC carried out explorations of the capabilities of mobile computers in an office setting which have resulted in the development of PARCTAB system prototype [11]. Narumi et al. [9] pose the technology requirements for ubiquitous environment and describe Fujitsu's real-time communicator ubiquitous middleware. Meanwhile, the FUSE (Future Ubiquitous Service Environment) group's focus is software architectures for ubiquitous service environments designing as well as privacy and context sensitivity [8]. Epp [5] stresses privacy and authenticated identity issues within ubiquitous environment. Burrell et al. [3] discuss the design and development of Semaphore, a contextually aware application for use in wireless networked environments.

2.2 Ubiquitous IS

As it was mentioned above, only few explorations have been done in this direction. However, we give a brief description of more or less corresponding works. "Docodemo-Display" is one of them and is a recently developed ubiquitous IS. It is a display-centric communication system, which is portable, interactive, and collaborative [10]. The goal of "Docodemo-Display" System is to increase the value of the place with helping to form local communities at that place. Another research is by Balke et al. [1]. In their work "A Situation-aware Mobile Traffic Information System" they present a prototype of traffic IS offering advanced personalized route planning, including added service like traffic jam alerting by means of SMS.

2.3 Factor analysis and Its Applications

Factor analysis is a branch of applied mathematics [6] and is generally referred to as a statistical tool for reduction of the number of variables and structure detection in the relationships between variables. First applied in psychology this method quickly became popular in other fields like economics, medicine, sociology, etc. as they often need to outline dominating factors out of huge volumes of variables they have to operate with.

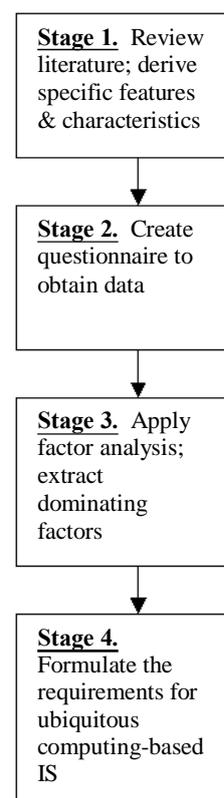
One of the conventional works using factor analysis describes the application of the latter to tests on Cheshire cheese [7]. The study reveals the relationships between a number of rheological experiments and essential subjective judgments on cheese. Another research applies factor analysis to psychological diagnostics [4] and particularly to the development of psychological tests and models such as the Cattell 16 Personal Factors model.

Numerous studies in various fields using factor analysis can be found. However, no studies seem to be carried out towards extracting the requirements for ubiquitous computing technology-based IS using factor analysis.

3. METHODOLOGY

Owing to well-developed concept we find factor analysis applicable to extracting essential requirements for ubiquitous computing-based IS. However, due to impossibility to collect data necessary for conducting empirical tests in a short period of time we describe only a methodology that comprises several stages (Figure 1). The first stage considers reviewing conventional literature in ubiquitous computing, services or environment fields and deriving specific features and characteristics. Next, on stage two we create a questionnaire for the experts in the related fields. The questionnaire will include the list of the derived features and characteristics that must be evaluated by the experts in terms of their significance. Having done this, we will get a data sample. The data will perform experts' estimations of the characteristics in scores. Having obtained the essential data we then apply factor analysis on stage three and extract dominating factors. Eventually, on stage four, relying on the results of factor analysis we formulate the requirements for ubiquitous computing-based IS.

Figure 1. Methodology of extracting the requirements to ubiquitous computing-based IS.



4. SUMMARY AND FUTURE WORK

The fact that very few works can be found on application of ubiquitous computing to IS and, therefore, the requirements for such an IS are not clearly articulated, makes this paper unique for it aims to

extract and classify those requirements. In this paper we delineate a general methodology of extracting the requirements using such a common statistical tool as factor analysis.

The future work within the research will lie in applying the described methodology to practice. Probably the most time-consuming stage will be collecting necessary data as it requires direct participation of respondents. However, having done this, it will be possible to acquire significant results which, in turn, will make a valuable contribution in ubiquitous computing-base IS development.

REFERENCES

- [1] Balke, W.-T, Kiebling, W., Unbehend, C., "A Situation-aware Mobile Traffic Information System", *proceedings of the 36th Hawaii International Conference on System Sciences*, 2002
- [2] Beadle, Peter H. W., Maguire Jr., G. Q., Smith, M. T., "Location Base Personal Mobile Computing and Communication", http://www.it.kth.se/~maguire/LocationAware/ieee_lan_98/, 1998
- [3] Burrell, J., Treadwell, P., Gay, Geri K., "Designing for Context: Usability in a Ubiquitous Environment", http://www.hci.cornell.edu/LabArticles/DesigningContext_Burrell.pdf
- [4] Chaplina, G., "Application of Factor Analysis in Psychological Diagnostics", <http://www.monax.ru/eleme.cgi?1889>, 2001 (in Russian)
- [5] Epp, E. C., "Relationship Management: Secure Collaboration in a Ubiquitous Environment", *Pervasive Computing Mobile and Ubiquitous Systems*, Apr-June 2003
- [6] Harman, H. H., *Modern Factor Analysis*, Chicago: The University of Chicago Press, 1976
- [7] Harper, R., Baron, M., "The Application of Factor Analysis to Tests on Cheese", *British Journal of Applied Physics*, Vol., 2, pp 35-41, 1951
- [8] Jansson, C. G., Lonnqvist, P., "Ubiquitous Service Environments", *ERCIM News*, No. 47, October 2001
- [9] Narumi, K., Yuhara, M., Matsumoto, T., "Technology Requirements for Ubiquitous Environment and Fujitsu's Realtime Communicator Solution", *Fujitsu Sci. Tech. J.*, 40, 1, pp 29-34, June 2004
- [10] Sekiguchi, T., Iguchi, S., Ishii, M., Maeoka, J., Ito, A., Furuhashi, T., "A Display Centric Ubiquitous Information System-"Docodemo-Display", *Symposium on Applications and Internet-WorkShops (SAINT 2004 Workshops)*, <http://csdl.computer.org/comp/proceedings/saint-w/2004/2050/00/20500709abs.htm>, January 2004
- [11] Schilit, B., Want, R., "The Xerox PARCTAB", <http://sandbox.parc.xerox.com/parctab/parctab.html>, 1995
- [12] Weiser, M., "Hot Topics: Ubiquitous Computing", *IEEE Computer*, October 1993
- [13] Weiser, M., "Some Computer Science Issues in Ubiquitous Computing", *Communications of the ACM*, July 1993
- [14] Weiser, M., "The Computer for 21st Century", <http://www.ubiq.com/hypertext/weiser/SciAmDraft3.html>
- [15] Yoo, K., Kwon, O., Suh, E., "UbiDSS: A Proactive Intelligent Decision Support System as an Expert System Deploying Ubiquitous Computing Technologies", 2004 (to be appeared)

IEEE 802.20 Based Broadband Railroad Digital Network – The Infrastructure for M-Commerce on the Train

Fumin Zou¹, Xinhua Jiang², Zhangxi Lin^{3*}

¹ Information College, Central-South University, Changsha 410075, China,

² Fujian University of Technology, Fuzhou, 350014, China,

¹xhj@mail.csu.edu.cn,²zlin@ba.ttu.edu,³fmzou@fjut.edu.cn

ABSTRACT

The broadband wireless access is emerging as a promising technology to meet the ever-increasing demand of M-commerce on the train. The traditional railroad communication system (RCS) is not only in charge of the traditional train scheduling, but also offers broadband WLAN services to passengers and provides the network platform to the intelligent railroad information system. This imposes a major challenge on the capability of the RCS in response to the increasing application requirements, particularly, the one for ubiquitous Internet access. To take the advantage of the rapid evolving mobile communication technology, this paper proposes an IEEE 802.20 based broadband railroad digital network, namely BRDN, for the next generation RCS. The paper further presents the scenario how BRDN operates and identifies the IP mobility as the major technical issue to be solved for BRDN. The predictability of mobile IP for the train-based applications will ease the difficulties in implementing BRDN. With the availability of IP standard for the next generation Internet - IPv6, BRDN will eventually become a reality.

Keywords: M-commerce; Broadband wireless network; Railroad Communication; Mobile IP

1. INTRODUCTION

With the advancement of information technology, M-commerce is undergoing a rapidly growth and exerting significant impacts on the economy. According to Finfacts' report based on several sources [8], the percentage of Internet users who use the wireless access grow from 16% in 2001 to 41.5% in 2004 (eMarketer, <http://www.emarketer.com/>). This figure will continue to get as high as 56.8% in 2007 with the total number Internet users almost tripled from that in 2001. In 2002, although only 2% of Internet users were using the "mobile cash" but 44% intended to use it later (ATKEARNEY, <http://www.atkearney.com/>). IDC (<http://www.idc.com/>) predicted that during 2001 to 2007, the M-commerce revenue in the US will grow from \$127 million to \$58.4 billion – more than 400 times.

Known as "wireless E-commerce", M-commerce generally refers to the use of wireless devices, such as mobile phones, to conduct electronic business transactions [17]. There are five value-added attributes of M-commerce to regular E-commerce [9][20]:

- Ubiquity: Available at any time and any location.
- Convenience: Convenient for users to operate.
- Instant Connectivity: Easily connect to the target.
- Personalization: Allows for preparation for personalized information.
- Location: Location-specific information and products.

As one of the most important scenarios of M-commerce, Internet access from trains has become an outstanding M-commerce issue among others. It is predicted that in the next five years, 625 million people will be traveling on Wi-Fi enabled trains around the world each year [3][5]. By 2008, train passengers will be spending 420 million dollars per year on in-transit wireless LAN hotspot services. Passengers in the train typically will stay a few hours to a few days in a small area that is like a mobile office more comfortable than staying in other moving vehicles. Thus, M-commerce allows passengers to conduct their business via the Internet. The key for the train-based M-commerce is the Internet access underpinned by mobile communications technology.

This paper is intended to propose the IEEE 802.20 based Broadband Railway Digital Network (BRDN), the wireless network infrastructure for the next generation railway communication system (RCS) based on the latest mobile broadband Internet access technology. The remaining part of this paper is organized as follows: Section 2 summarizes the status of RCS; Section 3 introduces the IEEE 802.20 standard; Section 4 presents a framework of BRDN based on IEEE 802.20; Section 5 discusses the key technical problems in implementing BRDN; finally, Section 6 concludes the paper.

2. THE EVOLUTION OF RAILROAD COMMUNICATION SYSTEM

* Primarily affiliated to Texas Tech University, USA

The traditional railroad communication system plays an important role in train distributing and programming, and inter-section wireless communication. At the same time, it is responsible for the railroad application data transmissions, such as train serial numbering, wind pressure remote sensing and railroad alarm messaging. As the intelligent railroad information system is highly regarded as the future value-added service on RCS, the limited data transmission capability of the traditional RCS is apparently the bottleneck for an advanced railroad information system. As ubiquitous Internet access becomes the trend, RCS should not only support the operation management of the railway system, but should also offer the Internet services to support M-commerce on the train.

The evolution of RCS relies on the advancement of the mobile communication technology (MCT). Historically, each new generation of MCT has refueled the RCS with the revolutionary improvement. As illustrated in Figure 1, MCT has led RCS's evolution with each innovative generation (the thin line represents the history of MCT, and the bold line indicates the following development step of RCS).

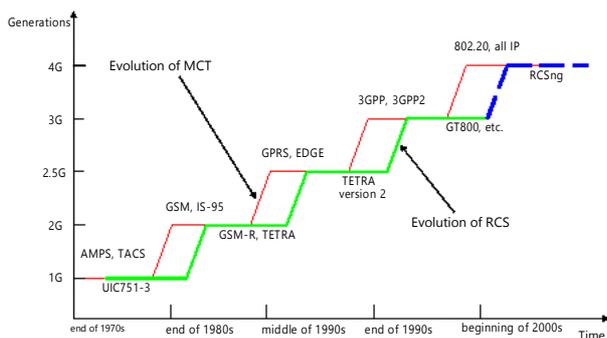


Figure 1. The development of MCT boosts the RCS's advance

In the last two decades, MCT has experienced rapid development, from the first generation (1G) to the third generation (3G) in technology, and achieved great commercial success. At the same time, RCS has evolved from analog communication system based on UIC751-3 protocol into the narrowband digital communication system based on 2G technology. The typical 2G technologies include GSM-R (UIC), iDEN (Motorola) and TETRA (ETSI), etc. The TETRA version 2 in planning by China's MoR possesses some 2.5G features [23]. Furthermore, the GT800 system developed by China's Huawei adopts the 3G technology (TD-SCDMA) and has been tested [10].

Now, the 4G technology based on IEEE 802.20 is rising on the horizon [2]. The basic idea of 4G is to merge the narrowband circuit-switching cellular phone technology with the broadband packet-switching technology. The IEEE 802.20 standard under discussion supports various vehicular mobility that classes up to 250 Km/h in a wide

area network environment with spectral efficiencies, low latency, high transmission rate, and QoS. Its goal is to enable worldwide deployment of affordable, ubiquitous, always-on and interoperable multi-vendor mobile broadband wireless access networks that meet the needs of business and residential end user markets. Although IEEE 802.20 demonstrates the promising future of mobile application, the narrow band digital mobile communication systems, such as GSM-R, TETRA, etc, are still the mostly available choice for RCS [7][21][24], only meeting the basic requirement for the operation of the railroad system. Even the 2.5G and 3G based RCSs, which are still under improvement, cannot satisfy the passengers' high demand for Internet access for M-commerce. This is mainly because it can only provide a peak bit rate up to 144Kbps when the mobile node is moving at a high speed. According to the status of RCS, it is a necessity to adopt the more advanced mobile communication technology.

3. IEEE 802.20 – THE KEY TECHNOLOGY FOR BRDN

Since July 1999, the IEEE 802.16 Working Group on Broadband Wireless Access (<http://grouper.ieee.org/groups/802/16/>) has been openly developing voluntary consensus standards for Wireless Metropolitan Area Networks with global applicability. Addressing the demand for broadband access to buildings, IEEE 802.16 provides solutions that are more economical than wired-line alternatives. The standards set the stage for a revolution in reliable, high-speed network access in the "last mile" of Internet by homes and enterprises [13]. On Decemberth, 2002, the IEEE Standards Board approved the establishment of IEEE 802.20 Mobile Broadband Wireless Access (MBWA) Working Group (<http://grouper.ieee.org/groups/802/20/>). It described the scope of IEEE 802.20 as:

Specification of physical and medium access control layers of an air interface for interoperable mobile broadband wireless access systems, operating in licensed bands below 3.5 GHz, optimized for IP-data transport, with peak data rates per user in excess of 1 Mbps. It supports various vehicular mobility classes up to 250 Km/h in a MAN environment and targets spectral efficiencies, sustained user data rates and numbers of active users that are all significantly higher than achieved by existing mobile systems.

According to the above scope, the basic features of IEEE 802.20 include compatibility, coexistence, distinct identity, technology feasibility, and economic feasibility [11].

According to the MBWA announcement, IEEE 802.20 is aimed at mobile communication, and its data rate can reach more than 2Mbps in high speed mobile application. IEEE 802.20 is the first real broadband wireless network standard that dedicatedly supports the mobility of network. A comparison between IEEE

802.20 and others mobile techniques for traditional RCS are shown in Table 1 [10][12][13][23].

Table 1 : IEEE 802.20 vs. Other Mobile Techniques used by Traditional RCS

Characteristics	GSM-R	TETRA version 2	GT800 (3G)	IEEE 802.20
Data rate	2.4-28.8Kbps	96-384Kbps	2Mbps, < 144Kbps in high speed	16Mbps, > 2Mbps at the speed of 250mk/h
Latency	About 1000ms	About 500ms	About 250ms	About 30ms
Spectral efficiency	200KHz/8channel	25KHz/4channel	About 0.2b/s/Hz/cell	> 1b/s/Hz/cell
Cell radius	5~10 Km	10~15 Km	2~5 Km	> 15 Km
Spectrum	Licensed bands 876-880/921-925MHz	Licensed bands 806-821/851-866Mhz	Licensed bands below 2.7GHz	Licensed bands below 3.5GHz
Switching method	Circuit	Circuit	Circuit/Packet	Packet

4. THE FRAMEWORK OF BRDN

In this section, we construct a framework for the BRDN underpinned by two wireless network standards - the IEEE 802.20 wireless WAN (WWAN) on ground and the IEEE 802.11x WLAN on train.

4.1 The Configuration

Figure 2 illustrates a configuration for a mobile network on the train implemented. A WLAN is deployed in a train with a typical configuration: The Ethernet (IEEE 802.3) cable is wired through all carriages via MVB (Multi Vehicle Bus); in each carriage, Wi-Fi access points are deployed with complete coverage of the area,

and are hooked up to the Ethernet; end users access this WLAN via their mobile devices, such as notebook computer, PDA, mobile IP phone, etc. An IEEE 802.20 client interfaced to the Ethernet via a router connects the WLAN to the base stations of IEEE 802.20 on the ground, which jointly offer the transparent services on the physical layer and data link layer between the Internet and the mobiles nodes.

Because of the availability of NAT (Network Address Translate) technology, the whole network on train may only need a single IP address. A DHCP server can automatically assign each mobile node of the network on the train a dynamic IP address.

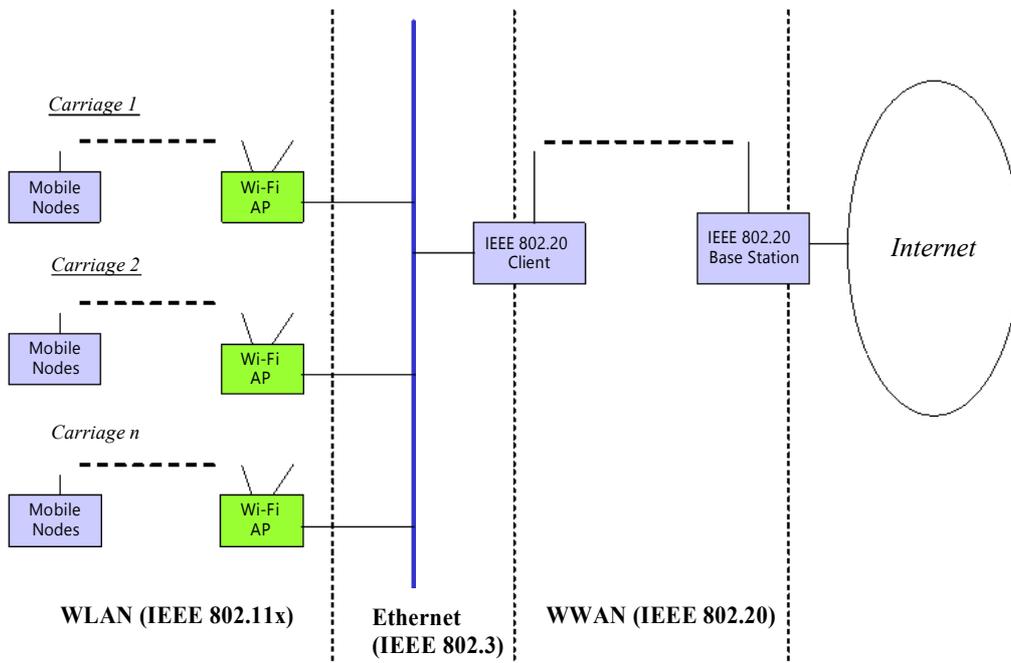


Figure 2. The Configuration of BRDN

The interconnection scheme of WWAN between the stations are still. As we discussed in Section 3, IEEE 802.20, IEEE 802.16x, existing cellular communication systems infrastructure, and the cable, because these base

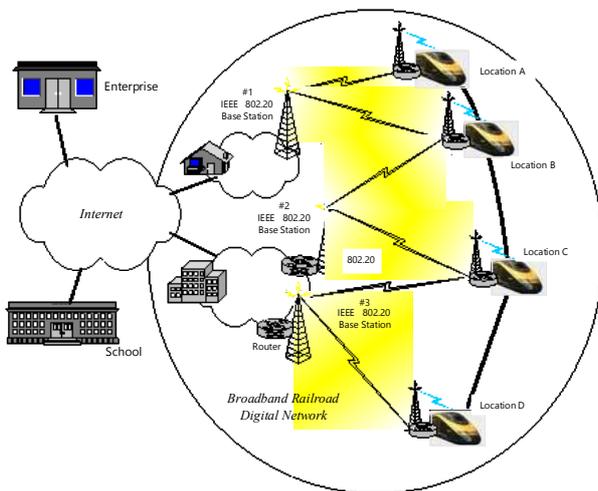
IEEE 802.20 is supposed to be the best choice for its high-speed mobility support, with high data rate, spectral efficiency and low latency.

The train-based IEEE 802.20 mobile client as a whole is on moving when the train is in operation, switching from one network to another over time (about 10-30 minutes at the regular moving speed). Section 5 will discuss the mobile IP problem - how the train-based client remains reachable from the ground stations, while moving around in the Internet quickly, regardless of its current point of attachment to the Internet.

According the above configuration, the proposed BRDN will adopt IEEE 802.20 and 802.11x standards to implement the function of physical layer and data link layer. The physical layer provides the air interface of wireless network, and the data link layer offers reliable data frame transmissions between neighboring hosts, by use of the bit transmission function provided by physical layer.

4.2 The Operation Scenario

The working scenario of BRDN is shown in Figure 3. When a train departs from location A, the #1 base station provides the Internet connection to the train via #1's network. When the train approaches location B, it will contact both #1 base station and #2 base station at the border because there is a coverage overlapping between the two base stations. After the IEEE 802.20 client on train switches the connection from #1 base station to #2 base station, #2's network starts to provide the path to the Internet, and the network connection handover is complete. The network handovers happens in the same way when the train arrives at location C and



location D.

Figure 3. The operation scenario of BRDN

5. MOBILE IPV6 - THE KEY ISSUE IN BRDN IMPLEMENTATION

Broadband wireless network has generally raised a number of research issues aiming at the requirements from the mobile applications. Although many have

proposed various resolutions for these issues, for example, mobile IP telephony, more effort has to be input to find the better ones that will eventually lead to commercialization. In the context of railroad based mobile applications, the same problems exist as in other mobile IP applications, but there is a possibility that the problems in railroad based mobile applications can be solved by alternative approaches with better outcomes because of the specialty of the application environment. This section briefly covers two most outstanding topics: IP mobility in BRDN, and VoIP for BRDN.

Mobile IP introduced a solution of bi-address, with a long-term IP address on a home network (HN) and a "care-of address" (COA) when away from its home network. This home address (HA) is administered in the same way as a "permanent" IP address is provided to a stationary host. The COA is associated with the mobile node (MN) and reflects the mobile node's current point of attachment. Every time when MN handover to a foreign network, mobile IP requires it to bind its COA to Home Agent (HA) [14][15]. If MN is moved far away to HA, the latency caused by handover will be long, and will lose the packets sent to MN, and the network performance will decline seriously. This Seamless Handover question has attracted wide attentions. Many researchers have put forward their proposal [1][16][18][23]. Much of them are based on mobility detection, and some are based on mobility predictive. But, because of the complexity of mobile node's moving, both the mobility detection and predictive are hardly implemented.

The IEEE 802.20 mobile client on train is always on moving along with train, switched from one network to another at a fixed trajectory, particularly, the period of handover can be predicted by the train schedule [22]. Based on these premises, BRDN adopts a Predictive Pre-Handover (PPH) algorithm. PPH is based on the highly predictive of train's moving from one network to the neighboring network. Not only the arriving time, but also the ID and router's address of the next network are predictable. It adopts many effective methods such as actively mobility detecting, pre-registering and pre-transmitting and storage. The flow of PPH algorithm illustrated as below:

- 1) Initialization: Create the entry for each network handover and get the predictive handover table according to the train's schedule.
- 2) Handover predict: BRDN predicts the next network's ID, router's address and arrive time t_p according to the handover table.
- 3) BRDN creates the frequency of detection and detects the foreign network actively.
- 4) Pre-registering: BRDN implements the preparations for registering, such as authorization and authentication.
- 5) Pre-transmitting and storage: According to the result of step 3 and t_p , BRDN copies the

communication data to the router of the network which about to handover.

- 6) BRDN triggers the handover process when it detected that the quality of the new network's signal is better than that of the former.
- 7) If there has any data missed, auto-transmit from the router's storage, and BRDN amends the predictive handover table according the result of this handover.

Internet, which can overcome the many known weaknesses of IPv4. With its 128 bits IP address IPv6 has realized the Stateless Address Auto-configuration (SAAC) [19]. Applied to the BRDN, IPv6 is able to configure each communication node on a train automatically with a unique global IP address freely. That implies that besides current IPv4 based dynamic IP assigning, IPv6 has the potential to provide a better solution. The operation for mobile IPv6 is shown in Figure 4.

IPv6 is presently the IP standard for the next generation

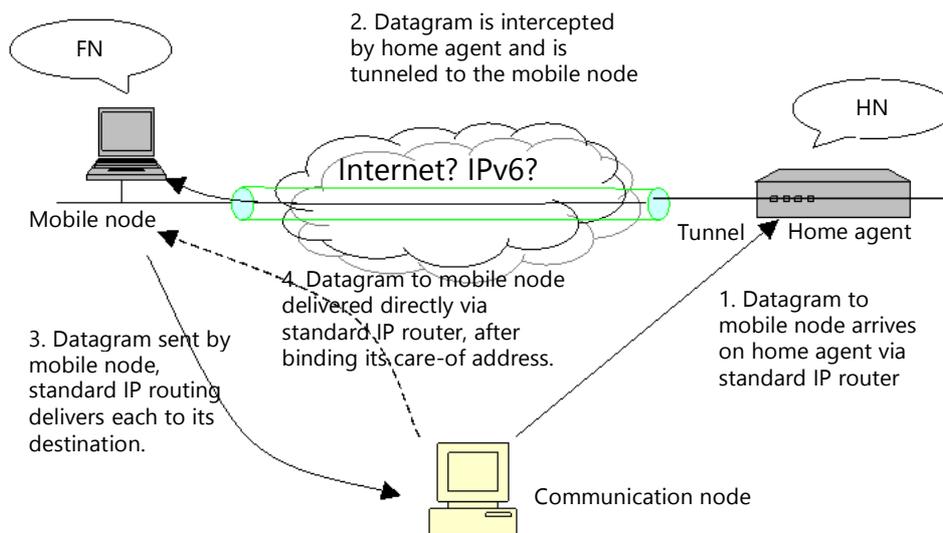


Figure 4. The operation of mobile IPv6

In a hybrid scheme, the WLAN on trains may still use mobile IPv4, which has been well explored recently, while the main connection to the Internet will be IPv6-based in order to support better IP mobility by taking the advantage of IPv6. A gateway between the WLAN and the IEEE 802.20 based WWAN will take care of the IP address translation.

Voice communication is a fundamental service demanded for RCS. In particularly, unless RBDN realizes the priority management, urgent call support, broadcast, multicast and quick establishment for call question, the RBDN based on IEEE802.20 can implement in rail-transit. VoIP application based on IEEE802.11x has been mature and deployed [4][6], which is the technology we introduced in the above sections.

5. SUMMARY

To extend M-commerce, the next generation RCS is a necessity for the advanced mobile network infrastructure. This paper presents current status of MCT and RCS, and introduces the latest mobile broadband wireless access technology IEEE 802.20. Although the standard of IEEE 802.20 is still under discussion and to be launched in 2005, its current specification has shown its promising potential in the

near future in supporting the ubiquity of M-commerce. We further compared IEEE 802.20's technology advantage with other RCS solutions, and identified that constructing the BRDN based on IEEE 802.20 standard is not only feasible, but also urgent. This paper further brought forward a configuration for RBDN, with the considerations in the required features of next generation RCS and the technology advantages of IEEE 802.20. We also discussed the mobile IP problem for BRDN – the key issue in implementing the next generation RCS, covering the predictive mobility management for mobile IP and VoIP over RBDN.

REFERENCES

- [1] Banerjee, N, W. Wu, K. Basu, and S. K. Das, "Analysis of SIP-based Mobility Management in 4G Wireless Networks," *Computer Communications*, Vol. 27, no. 8, May 2004, 697-707.
- [2] Bond, David, "4G - The Future of Mobile Applications," White paper of Fizmez, April 2004.
- [3] BWCS, "Railway WLAN Services," http://www.bwcs.com/brochures/VPN_brochure.pdf, 2003, last retrieved on October 22, 2004.
- [4] Choi, W Y, and S. K. Lee, "Maximum Number of VoIP Traffic Streams Served in IEEE 802.11i/e Wireless LANs," *IASTED International Conference on Communication, Network, and Information Security*, New York, NY, 10-12 Dec. 2003, pp.5-8.

- [5] Dorn, H., "Passenger Information Systems—Presence and Future," *Converter Technology & Electric Traction*, January 2004, pp.7-11.
- [6] Dutta, A, P. Agrawal, S. Das, M. Elaoud, D. Famolari, S. Madhani, and A. McAuley, B. Kim, P. Li, and M. Taulil, "Realizing Mobile Wireless Internet Telephony and Streaming Multimedia Testbed," *Computer Communications*, Vol. 27, no. 8, May 2004, pp.725-738.
- [7] Fan, Pingzhi, "Recent Advance of New Generation Railway Mobile Radio," *Journal of the China Railway*, Vol.19, No.2, April 1997, pp.102-108.
- [8] Finfacts Ireland, "Mobile E-Commerce," <http://www.finfacts.com/inttele.htm>, last retrieved on October 22, 2004.
- [9] Frolick, M. N., and Lei-da Chen, "Assessing M-commerce Opportunities," *Information System Management*, Spring 2004, pp.53-61.
- [10] Gong, Bian, and Xiaofu Ding, "GT800 Start an Era for Rail Transit Trunked Radio System," *Telecom World*, April 2004, pp.41-42.
- [11] IEEE 802.20 PD-03, "Mobile Broadband Wireless Access Systems," http://grouper.ieee.org/groups/802/20/P_Docs/IEEE%20802.20%20PD-03.pdf, November 2002, last retrieved on October 22, 2004.
- [12] IEEE 802.20 PD-04, "Introduction to IEEE 802.20 Technical and Procedural Orientation," http://www.ieee802.org/20/P_Docs/IEEE%20802.20%20PD-04.pdf, last retrieved on October 22, March 2003.
- [13] Intel, "IEEE 802.16 and WiMAX - Broadband Wireless Access for Everyone," http://www.intel.com/ebusiness/pdf/wireless/intel/80216_wimax.pdf, last retrieved on September 1, 2004.
- [14] Jiang, Xinhua, Fumin Zou, Yonghui Zhang, and Zhangxi Lin, "IPv6—A Boost for Implementing Mobile IP Applications," *Journal of Fujian University of Technology*, Vo.2, No.4, 2004, pp.379-389.
- [15] Johnson, D., C. Perkins, and J. Arkko, "Mobility Support in IPv6," RFC 3775, June 2004.
- [16] Kim, J H, S. K. Youm, S. H. Joung, and C. H. Kang, "Efficient Handoff Method for Supporting Micro-Mobility in Mobile IPv6," *IASTED International Conference on Wireless and Optical Communications*, Banff, Canada, 2-4 July 2003, pp.366-370.
- [17] Liang, Ting-Peng, and Chi-Ping Wei, "Introduction to the Special Issue: Mobile Commerce Applications," *International Journal of Electronic Commerce*, Spring 2004, pp.7-17.
- [18] Pei, Ke, Jiangdong Li, and Feng Guo, "Performance Search on Mobile IP Handoff," *Journal of XiDian University*, Vol. 29, No.3, June 2002, pp.328-332.
- [19] Perkins, C., "Mobility Support for IPv4," RFC 3344, August 2002.
- [20] Turban, E., D. King, J. Lee, M. Warkentin, and Chung, H. M. *Electronic Commerce: A Managerial Perspective*, Upper Saddle River, NJ: Prentice Hall, 2002.
- [21] Wang, Wenjing, and Zhangdui Zhong, "The Choice of Next Generation Railway Digital Mobile Communication System in China," *Chinese Railways*, No. 2, 2001, pp.34-36.
- [22] Zhang, Hongcheng, and Guangzhao Zhang, "Mobile IP predictive mobility management," *Journal of Circuits and Systems*, Vol.7, No.2, June 2002, pp.77-82.
- [23] Zhang, Yuwei, "The Analysis for TETRA and GSM-R," *Mobile Communications*, April 2002, 1-8.
- [24] Zheng, Jiang, "The Application of TETRA in High-speed Railway of Qin-Shen," *Mobile Communications*, May 2003, pp.80-84.

Individual Telecommunications Tariffs in Chinese Communities: History as a Mirror of the Future, and Relevance for Mobile Service Development in China

H. Chen, L-F Pau

Rotterdam School of Management, Erasmus University
F1-02, Burg. Oudlaan 50, P.O. BOX 1738, 3000 DR Rotterdam, the Netherlands
{hchen,lpau}@fbk.eur.nl

ABSTRACT

This paper addresses mobile service pricing and affordability issues in China. Taking history as a mirror, it shows individual tariffs existed at the dawn of telephony. A vision of future mobile services with individual tariffs is formalized which will suit specially well the culture of communities rooted in Chinese traditions. An analysis of current tariff conditions in China shows community-based individual tariffs will speed up the diffusion of mobile services to the majority of populations and benefit both economically and sociologically the development in China

Keywords: Individual Telecommunications Tariffs, Wireless Services, Wireless Communities, Telecommunication History

1. INTRODUCTION

The focus of this paper is on pricing and affordability of mobile services, addressing especially factors essential to a sustainable adoption of wireless communication in China. Taking a historical perspective around the world, it shows first that telecommunication services started in the 1880's with prices set for each user individually; an analysis of current situation in mobile communications especially in China is followed by a vista of the situation in approx. year 2030 (Section 2); it then presents (Section 3) a visionary version of mobile user pricing suiting especially well the Chinese culture of communities and the service dependencies therein (礼尚往来). The paper concludes in Section 4 with the suggestions that the adequacy of individual tariffs inside Chinese communities, further compounded with growing service creation skills, will fuel a fast improving affordability and use of mobile services.

The economic theory of pricing has traditionally been derived for physical goods, and from different angles:

1. either from a static equilibrium between supply and demand [1], including auction;
2. or taking in consideration ranked preferences for individual price formation [2];
3. or reflecting price dynamics with endogenous fluctuations due to market restructuring [3].

Most traditional telecommunication pricing schemes have been variants of the above, assuming limited capacity in either bandwidth or transmission capacity. These two assumptions have been made largely erroneous with the advent first of fiber optic transmission, and next of advanced radio coding/modulation/spectrum usage techniques.

Unfortunately, in general only little has been published around service pricing and affordability for end users,

which furthermore takes social networks into account, as in this paper.

2. THE HISTORY, PRESENT AND FUTURE OF WIRELESS SERVICE TRAFFIC AND AFFORDABILITY

This section is successively addressing the telephony usage and pricing in general in the 1880's, thereafter the current situation in China for wireless communications, and proposes finally a vision about wireless service pricing in e.g. the 2030's. (Date picked randomly to reflect the social and technology legacy and adoption curves.)

2.1 Users, Demand and Supply in the 1880's

At the dawn of the telephony services history, there were individual tariffs! The telephone was patented by Alexander Graham Bell in 1876. For quite some years, and in some markets, it served mainly some high level civil servants and privileged people (trade, news), and was the symbol of wealth and social rank. Telephone subscribers were not designated by a number, but by their name; picking up the phone would get the operator, who would then ring and connect the desired party by a polite support staff protocol.

As a luxury service at that time, the demand for telephony was limited (lack of network capacity effect). Phone calls were mainly local, with national and international calls a rare phenomenon except in some key cities. The demand patterns by called party, between office and home, or between national parties and international parties, were by and large not analyzed, as their influence on telephone company tariffs was minimal. Each individual was eventually satisfied, or not, in a binary way with whatever the telephone companies offered (since the person was rich or the services were paid by the government). The individual

demands were price insensitive, which resulted in an inelastic demand curve as shown in Figure 1.

Regarding the supply of telephony services, there were in general only geographically limited pure monopolies, and the capacity in each country was dominated by two or three local companies able to invest in interconnections.

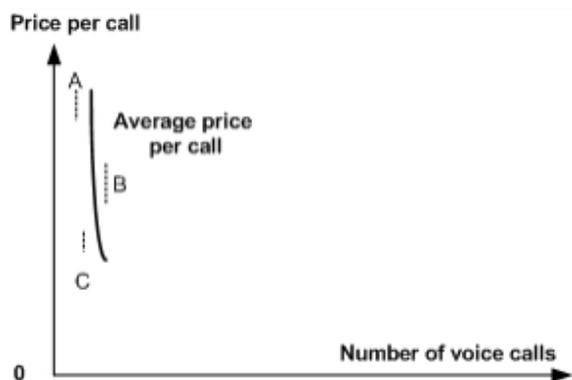


Figure 1: Service demands at the dawn of telephony. A, B, C represents the individual demands, which were limited and price insensitive. The average telephony demand curve at the dawn of telephony is also shown in the bold curve, which is inelastic. All voice calls here are assumed for the same destination/distance.

From the very beginning, although the reduced sets of telephone users and suppliers were restricted, charging patterns diverged very early between flat rates and individual usage-based rates. Numbers of call attempts, physical destination and the duration were manually recorded – but the pricing of the calls was a matter between the telephone company sales person and the service customer (who was then not a subscriber); usage, rank, fame, location were all taken into account, and the settlement was done by a bank note or cash. In some countries, an annual contract fee was charged, while others started with flat rates for a fixed number of calls (not even based on duration or distance).

Case 1: In Sweden, year 1881, the city of Gävle called for bids to supply a local telephone system. The Bell Company in Stockholm offered to install and operate a system for SEK 200 per user per year (2004 value = SEK 9102)¹, based on a minimum of fifty five-year user contracts. The local enterprise Ericsson offered SEK 275 per user (2004 value = SEK 12514) and thereafter to operate it for SEK 56 (2004 value = SEK 2458) per user per year². The rate of annual operation cost, to Gross National Product (GNP) of SEK 310/capita in 1881³, is 18%.

¹ "Consumer Price Index," Statistics Sweden, 2004.

² "Lars Magnus Ericsson - A brief biography," Telefonaktiebolaget LM Ericsson. Available: http://www.ericsson.com/about/compfacts/history/lars_magnus_ericsson/company_born.shtml.

³ "The gross domestic product of Sweden and its composition 1861-1955.", Östen Johansson.

Case 2: In Los Angeles (U.S.A), in 1888 the common policy of the telephone company was high-quality service and high costs. "The minimum flat rate in central Los Angeles in 1888, for example, was USD 4 per month (2004 value = USD 80.37)⁴ plus two cents a connection after the fortieth call (2004 value = USD 0.40). This rate equaled about 10% of the average nonfarm employee's wages" [4]. At the time, the penetration rate of telephone in the U.S is 1.08%⁵.

In all cases, the supply was scarce and inelastic to prices, as there was no choice, nor competitive mechanisms. And, although prices were fixed individually or bundled into packages with fixed numbers of calls, they were extremely high, which resulted in the exclusion of the majority of population from accessing the services due to unaffordability.

2.2 Current Situation for Mobile Communications and Especially in Relation to China

The worldwide GSM mobile phone users reached 1 billion in February 2004, almost one sixth of the world's population⁶; besides other mobile communications standards also have user and subscriber bases adding up to that number, for a worldwide grand total of approximately 1.4 Billion (1,340,667.7) at the end of 2003⁷.

In China, the first public mobile operator (China Telecom) started its services in 1987 under heavy government involvement and as a state-owned enterprise, with 700 subscribers. With a fierce growth, in October 2003, the number of mobile users exceeds the number of fixed telephone users⁸. By the end of 2003, the overall penetration rate among the total population was approximately 20.9%. There were approximately 290.31 million mobile users at the end of March 2004, with a growth of 60 millions of users per year, according to MII statistics⁹.

In China, mobile tariffs, as part of the telecommunication tariffs, are guided by "Regulations of the People's Republic of China on Telecommunications"[5]. A Telecommunications Law is supposed to be promulgated in 2005. Tariffs are divided into three categories:

1. Government-set tariffs;

⁴ "Consumer Price Indexes," Bureau of Labor Statistics, U.S. Department of Labor, 2004.

⁵ "Penetration rate of Consumer Technologies (1876-present)", FCC, U.S.A

⁶ "GSM: the business of a billion people." Cannes: GSM Association Press Release, 2004.

⁷ "World Telecommunication Indicators - Basic indicators," ITU, 2003. Available: http://www.itu.int/ITU-D/ict/statistics/at_glance/basic03.pdf

⁸ "Telecommunication Industry Monthly Statistics, October 2003," Ministry of Information Industry, China.

⁹ "Telecommunication Industry Indicator, 1st quarter 2004," Ministry of Information Industry, China.

2. Government-guided tariffs;
3. Market-based tariffs.

Basic mobile services (voice call, roaming, and long distance call) can follow either government-set tariffs, government-guided tariffs or market-based tariffs. Value-added mobile services (call forward, line identification, SMS, etc) can follow government-guided tariffs or market-based tariffs. When there is intensive competition, the tariffs can be set based on the market mechanism. "State-aid" operators are entitled to set price fluctuating within 10% of the government-set tariffs. There are two dominant mobile operators (China Mobile and China Unicom), plus two minor ones.

The current mobile tariffs are based on a regulation issued in 1994 by the former Ministry of Posts and Telecommunications. Postpaid users pay a monthly subscription fee of CNY 50 and the tariff is CNY 0.40/minute for local calls, CNY 0.60/minute when roaming. Prepaid users do not pay the monthly subscription fee, and the corresponding tariffs are CNY 0.60/minute for local calls and CNY 0.80/minute when roaming.

Table1: Ratio of ARPU to disposable income.

	Urban usage	Rural usage
Weighted ARPU (CNY)/month	129.93	56.38
Disposable income (CNY/year)	8472	2622
(Weighted ARPU)*12/ Disposable income	18.40%	26%

We approximate urban users to be the postpaid users, and rural users to be the prepaid users. The weighted ARPU is the average revenue per user, weighted within each population category (urban/postpaid, rural/prepaid) by weight equal to market share applied to published ARPU of two large mobile operators: China mobile and China Unicom.

Due to high income disparity in China today, the ratio of the ARPU¹⁰ to disposable income urban and rural¹¹ (see table 1) reminds closely of the situation in the 1880's described above in Section 2.1, with a ratio of 18.40% and 26% respectively. The overall average ARPU is CNY 76.3/month (2003). The ratio of the overall average mobile voice usage cost per year, to Chinese GDP per inhabitant CNY 9025 (2003) is of 10.14 %. Whereas the overall mobile penetration rate in China among the total population was approximately 20.9% at the end of 2003, with high usage subscribers representing an estimated 20% of the subscriber base, or

¹⁰ The ARPU data are acquired from the 2003 annual results of China Mobile (Hong Kong) limited and 2003 annual results of China Unicom Limited.

¹¹ "Statistical Communiqué on the 2003 National Economic and Social development," National Bureau of Statistics of China.

4.2% of the population. For all practical matters, high volume users (privates or enterprises) still largely negotiate their rates. Meanwhile, the majority of the population (79%) still can not afford the expense on wireless communications, reminding of the situation depicted in Section 2.1.

Although the regulatory situation has obviously evolved rapidly, in purely economic terms, demand and supply are not yet happening on an equal-access basis and the suppliers are closer to an oligopolistic behavior than to a pure competitive one, partly due to the government directives.

2.3 A Vision of the Tariffing Situation in Approx. Year 2030 ("The Bright Future")

Twenty years from now on, advanced personal communication technologies enable people to stay connected anytime, anywhere with access network alternatives. Users' devices can seamlessly roam between PAN, WLAN, WAN, with new services yet unknown. With a penetration rate of wireless devices of over 100 percent over the entire population in most countries, how mobile services are provided will be quite different from now.

While connectivity is the most cherished property as in year 2000 (see [6]), the key values of mobile service in the 2030's will be totally different. By a combination of a large number of users, technology improvements, and operator productivity gains, the pure transport and access tariffs for wireless will have plummeted to very low values. Content-based service will generate certain revenue. Content exists in two variants:

1. "static", from data warehouses with only slow modifications;
2. "dynamic", from real-time information sensors and other sources, including user-originated content.

By profiling and data mining, besides personalization selections, the service operators will know in real time much more about users than now. However, in an age with information "overload", static content has no commercial value except for those with copyright of the creators, or information access provisions. Advertisement will generate some revenue for service providers, but it will probably not be sufficient to support a whole mobile communication industry living from the "law of large numbers" and from very low tariffs. User driven service personalization will enhance some tariffs, but the cost of provisioning of such services will offset revenue gains.

Mirroring the early trend in video/broadcasting industries for dynamic content, the relative share in tariff bundles of the intellectual property right owner will be larger; the multiplication of dynamic content channels will add to tariff bundles price pressure.

Consequently by say 2030, the true value of wireless services lies in the interactions, where the services are formed as a result of multiparty interactions. These interactions can be divided into three categories:

1. Community based human-to-human interactions;
2. Cluster based machine-to-machine interactions;
3. Human-to-machine interactions.

In this paper, we focus on the first category. Sociologist Barry Wellman [7] had defined communities as: "Networks of interpersonal ties that provide sociability, support, information, a sense of belonging, and social identity." We assume that all the members of such a community rely amongst other communication mechanisms, on the usage of mobile technology in the 2030's.

Because of their business, sociological or process-linked nature, the communities mentioned above rely on a membership fee (in kind or in money, eventually free) and on managed access privileges, akin operators customer care administration, but in diverse forms and with more freedom in the organization thereof. More precisely, belonging to community requires an identity and membership for a user in order to receive services. The community provides collectively information, know-how, and services to members through interactions and access to static or dynamic content. The access to the services inside a community is based on prices and based on competition, or possibly quasi real-time spot service prices. There are also prices for community peering arrangements, or community-to-community interactions, but these may be on a flat basis once a service-level agreement has been enacted between them. One key driver of these communities is obviously the affordability in economic terms, and the obligation in social terms, to join one or several communities. Section 3 explains in more detail the cost the revenue formation justifying this affordability.

The above analysis leads directly to pointing at the future potential of individual tariffs between now and the 2030's, although these would obviously be provided in a completely different way to the 1880's in terms of user population, technologies and management processes.

In China, the traditional "families" and groups culture (人以群分) leads to more diverse communities. In addition to business, occupations and hobbies, communities are often formed under a same origin, or under a same dialect [8]. The above pricing settlement arrangements suit especially well the various communities. Take the example of a family-based community: the flat rate will be very low (close to zero), similarly within a rural village. But the differences lie in the fact that in a formally organized community, "such as in a village", affordability will be higher as set

largely by users themselves in view of their social, economic and information needs, which may result in non-zero membership fees.

3. INDIVIDUAL TARIFFS

3.1 Definition of Individual Tariffs for a Set of Communication Services

"*Individual tariff*" means that each individual sets a tariff for himself/herself for a specific set of services provided by the community, whether this service is user-defined or community-defined. Even if that individual belongs, say to an enterprise, the members of the enterprise may have different individual tariffs, simply because their service demand and content flow (contributions and receipts) are different. Even, different users of the identical service (if any such exists) may value and price it differently as they decide to belong to different communities as their choices.

In the definition above each community ends up supplying a number of services of which only some are initiated by consensus at community level or by the service manager. A community does not need to own part or the entire fixed and wireless transmission infrastructure, which can be sourced competitively from an infrastructure owner.

The above definition says nothing about the service provisioning duration the community commits to its members, or which communities commit to between themselves. Duration of the service will be one attribute in the multi-attribute service demand from the user in a community; e.g. sporadic uses are possible, just as are long term ones, but the difference with today is that they can be different and user-specific.

The above definition says nothing either about the transparency of prices and pricing provisions. More precisely, the information disclosure rules are of three types at least, with one between the member and his community (especially if this is managed formally), one within a community, and the last one between communities. These information disclosure rules may lead to price equalization but this effect is twice limited in that it can only happen for the same service, while each community member will have a different service profile.

Very importantly, per this definition, when the end user requests a certain service from a community, he/she is also responsible for the existence and survival of the community through contributions (money, but also information, knowledge). The individual will take and share the risk if the service is underfunded and ceases to exist. So, if the user appreciates the service, he/she may end up paying slightly more than other members, or even than users of the same service in other communities, to ensure the existence of the service. The

person can also consider paying an insurance premium to the community to cover against the risk of service loss.

By paying an individually negotiated premium, which can be a very small amount of money, the user maintains his access to services, but also by mutualising this risk, he contributes to the survival of the community [9].

There are several countries where some operators are on an experimental basis, or publicly, having individual tariffs which are not just options inside or outside a wireless “subscription plan/bundle”: Estonia, Norway and South Africa. In first two cases, it is part of business model and last one as part of policy to balance with disposable income (“universal service provisioning”).

3.2 Business Model of a Community

The income of a community will be made up of a combination of:

1. the member fee (once per year);
2. the competitive specific service usage revenues within the community;
3. service usage revenues generated from non-members;
4. the possible premium income from members who select to cover themselves against specific service disruption;
5. flat fees from other communities;
6. and, last but not least, contributions in kind (work, information, know-how, knowledge, innovation) by members.

On the cost side, they include:

1. costs of managing community memberships; but this cost does not bear marketing and publicity costs, so it will be far less than with today’s public operators;
2. investments in infrastructures and possibly service access devices;
3. partial service creation expenses, possibly shared with community users or other communities;
4. service provisioning and operations expenses, with in some cases community members being member-employees;
5. flat fees to other communities;
6. and, last but not least, intellectual property right expense from service creation and innovation, and from access to information or knowledge shared in the community.

We estimate that the community management overhead share will be 25%, the community service creation share 60% and the transport plus access share (to infrastructure owner) will be 15% of the total costs.

3.3 The Demand Analysis

The total demand will be large, as people can

personalize their services according to their needs, and furthermore because community proliferation may multiply the effect. Compared to the situation with generic public services in limited number, used by very many, communities offer viable alternatives in terms of revenue and demand. The reason is that within a community, the above two multiplicity factors are replaced by four:

1. number of community members;
2. overlapped sub-communities that share some common interests but have some interests and service demands different from others;
3. number of specific services for a sub-community;
4. number of common services to all members in the community.

These four distributions allow to replace flat rates for all, by individual tariffs, subject to the condition that each community has detailed network management and monitoring tools to be able in real-time as well over membership durations to quantify equilibrium break-even tariffs (see [10]). People will pay the amount of money, and put in their individual contributions, exactly according to how they value the personalized service bundles offered or requested by them. This results in price insensitive individual demands. Although for each specific service, the preference from each user will be different, but the average demand across large number of user again leads to a price inelastic demand curve (see Figure 2); in this way, we are back in history (see Figure 1). However, the new demand curve is much more to the right of in absolute value in the “number of service requests” axis and much lower in the price axis.

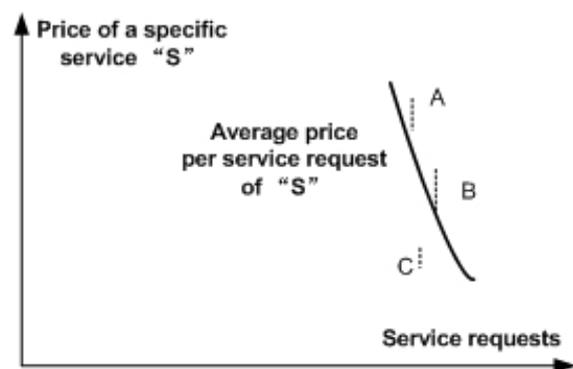


Figure 2: Service demand for one specific service “S” at 2030’s.

A, B and C represent individual A’s service profile: (S, A1, A2, etc), individual B’s service profile: (S, B1, B2, etc) and individual C’s service profile: (S, C1, C2, etc).

The average demand for service “S” is shown in the bold curve. Note there is a “service proliferation effect” driving the curve slope.

Case 3: A good example of community services are from the largest Chinese game community “Lian Zhong” (联众世界 www.ourgame.com). It offers

online and mobile gaming, which include board games, card games and arcade games. It has over 130 million members, including 1 million paid members. Member fees are CNY 35 (90 days) or CNY 120 (1 year). However, membership days can be traded into credits which can be spent in the games. (E.g. 10 days of membership for 300 credits.) Paid members can, among other services, buy credits, have privilege joining and selecting the games, set up their sub-communities and can enter the "hall of fame" based on their scores and credits. Normal members can access common game services, with limited capability; they can not buy credits but have to earn them in the games. There are over 3600 sub-communities which are formed under different basis. An average sub-community has 600 members, and over 20% are paid members. Besides the common services, members in each sub-community can have their own specific services. Generally, each sub-community offers 2-3 specific services, like monthly training or tournaments but tailored to their favorite games and within a closed community. Currently, the access to services offered by sub-communities is free.

3.4 The Supply Analysis

The supply of services will be abundant, and the price will be low, driven not only by the deregulation of telecommunication services and technology advances, but above all by the freedom to define, request and bear a share of the risk around service creation. In the Chinese case, and maybe others, additional factors are the traditions within a community and their closed nature.

Still backbone transmission will be essential, as will be different authentication/roaming/settlement functions, but it remains to be seen if traditional operators born in mindset in the 1880's (analog voice) or 2000's (wireless) can offer competitive services in flexibility, price, quality and scalability to what some larger communities may do themselves. Do not forget that due to exploding traffic demand, and competition driven by many community-clients, the revenue from pure transport or access will become minimal compared to the added value of personalized services.

4. CONCLUSION: INDIVIDUAL TARIFFS IN CHINA?

The above analysis and evolution is highly relevant to the cultural, technical and sociological evolution in China –it shows a pattern in China today as to mobile adoption and affordability similar to the early days in the 1880's for fixed telephony on the user side.

A graduate adoption of community-based individual tariffs, first on connectivity, then on content, and finally

on interactions, will speed up the diffusion of mobile services for the majority of population in China. While various demands are met individually by user-set affordability, China can avoid the competitive model between large operators offering reduced set of standardized services and at the same time obtain a sustainable development in economic aspect.

One the other hand, the importance of individual and community pricing scheme is furthermore compounded by the importance in China of regional dialect and regional traditions as well as community demographics.

Community-based individual tariffs can be fueled by the very fast adoption of new technologies (WLAN etc.), and cultural linked services (e.g. mobile gaming). Furthermore, the skills base in software for service creation is a pre-requisite, the very presence of which in China may further accelerate this evolution by community based service needs.

However, the vision of individual tariffs envisioned here is not implemented in China, nor yet supported by regulatory policies. A balance has to be found between the emergence of communities, government policies and the opening up of Chinese communication market.

REFERENCES

- [1] P.A. Samuelson, W.D. Nordhaus in *Economics*, 16ed, New York: Irwin/McGraw-Hill, 1998, pp. 51-57.
- [2] F.A. Hayek, *Individualism and economic order*. Chicago: University of Chicago Press, 1980.
- [3] J. Tuinstra, *Price dynamics in equilibrium models: the search for equilibrium and the emergence of endogenous fluctuation*. Boston: Kluwer Academic Publ., 2001.
- [4] C. S. Fischer, in *America calling: a social history of the telephone to 1940*. Berkeley: University of California Press, cop. 1992, pp. 38-39.
- [5] *Regulations of the People's Republic of China on Telecommunications*, Decree No. 291 of the State Council of the People's Republic of China, September 20, 2000.
- [6] A. M. Odlyzko. (2001, February). "Content is Not King". *First Monday* [online journal]. vol. 6. Available: http://firstmonday.org/issues/issue6_2/odlyzko/index.html
- [7] B. Wellman, "Physical Place and Cyberplace: The Rise of Personalized Networking," *International Journal of Urban and Regional Research*, vol. 25(2), pp. 227-252, 2001.
- [8] R. Kluver, J H. Powers, *Civic Discourse, Civil Society, and Chinese Communities*. Stamford: Ablex Publishing, 1999.
- [9] S. Harrington. G.R. Niehaus in *Risk Management & Insurance*, 2ed, New York: McGraw-Hill, 2003, pp. 54-74.
- [10] L-F. Pau, "The communications and information economy: issues, tariffs and economics research areas," *Journal of Economic Dynamics and Control*, vol. 26, pp. 1651-1675, August 2002.

Managing Critical Incidents in the Fine-Paper Supply Chain through Mobile Collaborative System

Vaida Kadyte

Åbo Akademi University, Turku Centre for Computer Science, IAMSR
Lemminkäisenkatu 14 B, Turku 20520, Finland
vkadyte@abo.fi

ABSTRACT

Most of the CSCW innovations related to mobile technologies are at the research stage and have been launched commercially, as they are developed within labs and lack direct practical relevance. This paper reports the findings of a case study on real-world companies operating in the fine-paper supply chain and places emphasis on the requirements for supporting mobility within specific collaborative settings. A mobile web-based system for managing customer complaints in a collaborative group work environment is described, which we have designed and developed for the fine-paper industrial setting. The benefits - including stronger customer relationships, lower operating costs and better use of human resources - are available to any complex customer care activity that follows activity-centered collaboration and seizes the initiative via a web-based system with mobile access.

Keywords: collaboration, mobile system, complaint handling, paper industry

1. INTRODUCTION

The ever-increasing spread of mobile communications, rapid development of sophisticated smart phones and forthcoming future of IP convergence make it possible to adapt computer-supported cooperative work (CSCW) applications to mobile devices. Our motive for this study is twofold. We are witnessing the emergence of networked organizations that are based on flexible business processes and rely upon the dynamic work of teams that temporally coalesce. Increasing business velocity and mobility require collaborative applications and systems in B2B settings. One of the most time-sensitive fields in every business operation is internal and external customer service. Therefore major software providers, vendors and consulting companies offer various mobile deployment options for their CRM products [16]. Despite the contribution of such systems to supporting remote activities, these solutions are unwieldy and mobility in collaboration is often overlooked. Secondly, we aim to address an issue that has recently emerged within CSCW and still received relatively little attention – the requirements and conditions to support mobility in B2B collaborative activities. A growing number of studies exist on mobility in collaboration that offer findings about the actualities of mobile work in general and characterize the problems of mobile work [1, 11, 6, 15]. However, only a few of them have presented actual prototypes for mobile CSCW, featured them in empirical studies of work and communication [13, 3, 7, 12] and provided enough evidence of the implications for supporting mobility and enhancing collaborative work in corporate settings. As a number of authors have commented, activity-centered research on collaborative work is now beginning to emerge [8, 10]. This paper is based on the latter concept and presents a mobile collaborative system that places a strong emphasis on the requirements for supporting

mobility within specific collaborative settings. We report on studies of communication and collaboration in a vertical corporate market of the fine-paper industry, and discuss the implications of the horizontal mobile solutions for supporting collaborative work in the complaints handling process and more generally enhancing customer relationships.

2. CRITICAL INCIDENTS IN THE FINE-PAPER SUPPLY CHAIN

The case study reported here centers on three real-world companies, participating in the supply chain for high-quality fine paper. It aims to reveal the general condition of customer care that are specific to the fine-paper industry and focus on what is needed to improve the handling of complaints via a CSCW system. The fine-paper business within the pulp & paper industry produces a multitude of paper-based products. The industry is complex and depends on good customer relations with a great variety of players, ranging from paper mills and printers to retailers and consumers. A successful commercial publication is a sum of elements: content, layout, illustrations, font and, often very crucially, the paper and print quality. Despite the complexity, there are some key elements to becoming more successful on the market. In the fine-paper supply chain you need your customers' business. Most customers today want more flexible service concepts and more transparency in the supply chain and are willing to work more closely in production planning, logistics, etc. to get that. In the fine-paper industry, there is a lot of talk about customer service focus within manufacturing settings, but surprisingly few companies have envisioned where to start from and how to accomplish this. Our case is no exception, and critical incidents are often overlooked. Critical incidents related to fine-paper products consist of feedback and claims, which together

are referred to as complaints. There is no room for setting up a call center, since to resolve the conflict requires group decision-making and follows complex checkup procedures on the part of the technical customer service (TCS) working at the mill and of the personnel maintaining the printing machines at the customer site. Besides being a complex activity that requires group decision-making, coordination and information sharing, handling complaints related to fine-paper products is also a seasonal activity, which puts additional pressure from time to time on both customers and TCS. Another factor is related to the timeliness of the product to the end user, i.e. the consumer. Moreover, to keep abreast of market demands, printing houses aim to incorporate state-of-the-art technology. In the era of the mobile internet, this means customers in the fine-paper industry will be equipped with modern and efficient terminals, will submit their complaints and questions to the paper mills and expect to be answered promptly. In concluding this section, we stress that handling complaints in the fine-paper supply chain is a complex, group-work activity that under seasonal time pressure requires a high degree of collaboration.

3. RESOLVING CRITICAL INCIDENTS WITH TRADITIONAL IS TECHNOLOGY

Design, development and implementation of collaborative systems imply an understanding of group processes, i.e. the essence of collaborative work. Before describing a suggested collaborative system for complaint handling, let us have a look at the existing communication and work structures of complaint handling.

3.1 Customer Service Tools for Handling Complaints

In traditional markets such as fine-paper products, customer complaints are considered an important source of information regarding areas of customer satisfaction and profitability. A recent study that examined complaints data [4] found that more customer dissatisfaction was generated by problems with customer service centers than by product failure. However, the first difficulty arises in the context of the fine-paper industry: what communication channels are most beneficial when trying to serve customers fast, individually and successfully? In the past few years online customer service has been receiving more attention, as the Internet has become a powerful tool for strengthening relationships between firms and their customers [17]. Recently, one of the paper mills' priorities has been to provide more online service for corporate customers. The paper-producing company has upgraded and improved its website. The new site offers a frequently asked questions (FAQ) section, and interactive discussion forum and a promotional goods section with marketing materials online. The FAQ page of an Internet producer's site provides answers to questions that are asked by customers on a regular basis. The disadvantage, however,

is just as clear: auto responders are standardized and can therefore not be called a customized service. When customers ask questions they will not be satisfied with a "canned" answer. Especially in complex processes, business customers want to communicate with humans. Auto responders can never be a substitute for a personalized answer and can therefore only be a subordinate part of a functioning customer care system.

While some researchers identify e-mail as one of the most pervasive collaborative technologies in organizations [18], others regard it as one of the worst communication media because of information overload and the interruption to work [9]. Even if the former can be proved, it does not seem to be an appropriate solution for handling complaints in the fine-paper supply chain. Failure to treat customer questions fast will result in dissatisfied customers and poor perception of the company. A rather popular customer care tool in the paper industry is an auto e-mail responder which is widely used especially when the technical customer service personnel is not available on line. The auto responder sends out a standardized reply to every incoming e-mail before any human looks at it. This means that one aspect of the customer relationship concept, instant customer care, is fulfilled. However in most cases the customer realizes that the service center on the supplier's side is a "dead end" providing usual apologies for a personnel being out of the office. Despite the modernization efforts and introduction of online customer service tools, there is a strong preference for face-to-face (F-to-F) interaction channels among technical service employees and management. "When we see people face to face, we really invest in the relationship", says a technical service manager at a paper mill. These moments of physical co-presence and F-to-F conversation are crucial to patterns of social life, but emphasis on meeting and thus the related delay in response time arising out of different modes and form of travel, as well as related traveling expenses may negatively contribute to the customer relationship. Many alternatives to traditional face-to-face interactions now exist without the need to travel to a common location. Mobile technology is the most efficient communication medium so far: it is real time, always and everywhere with the user, fast and easy to use, and customer service can indeed be reinforced by the mobile use of Web technology. When it comes to collaboration, 'togetherness' implies that group work takes place through the same artifacts. For certain processes in the supply chain, such as handling complaints, the cooperation takes place mainly through complex communication and understanding. This suggests that an interactive and efficient medium that enables instant information delivery, easy access to that information flow and sharing among participants should be considered an appropriate communication artifact for resolving industrial complaints. However, the picture becomes even more complicated in our case, as both tasks of complaint handling and individual difference

factors affect the way groups use the medium of their choice available to them.

3.2 The Study

In the case study reported here we have focused on developing new IS solutions and used constructive research methods based on action research methodology. Multiple data sources provided more accurate evidence of the complexity of the process. In our data collection effort, we used F-to-F interviews and documentary materials as the primary source of data. We conducted private interviews with each of the 10 focus group members from the paper producing company, customer printing houses and wholesaler. The official documents we received (a standard flow diagram and summaries of complaints handled in the year 2003) gave us a picture of how the complaints had been handled by the paper-producing company. As an independent external facilitator of the process, the author of this study conducted a series of expert surveys to enable customers, service technicians and management at the mill to provide data for defining problems and opportunity situations regarding customer service. Responses were obtained from 16 experts. Fragmentation of working time, information flow and working space were used as measures to uncover shortcomings within the existing process and potential areas for improvement, where combination of mobile and web technology served as an enabler. Data collecting process, covering the stages of feasibility and development through to preliminary experiment, took a period of one and a half years.

3.3 Dimensions of Work and Resulting Requirements

One of the motives for using collaboration systems is that CSCW is more effective and efficient than traditional forms of teamwork. A group of team members has more information and possesses more knowledge than any individual member alone. Therefore groups are better than individuals at understanding problems, in noticing errors; as a result, more alternatives are generated for problem solving. A key concept of Systems Theory [4] can be applied in the complex industry of fine-paper products: the effectiveness of group work in resolving complaints and the quality of the corrective actions suggested can be larger than the sum of what is produced by an independent individual. But incomplete access to information can hamper successful completion of the group work. The group process can be time-consuming, costly and might even jeopardize good customer relations without appropriate coordination to integrate participants' contribution in solving critical incidents. Co-location, temporality and contextuality are not mentioned in the literature as variables of groupness, but they are one of the most essential characteristics of individual work. If not taken seriously enough, these dimensions of individual work may constitute barriers to group collaboration.

Spatiality. The existing complaints handling process suffers from coordination difficulties when the process actors are working remotely. These people are located at different offices across the continent and use various technological artifacts to communicate. But not all the actors have the same intensity and scope of mobility. Technical managers at the paper-producing company are key decision-makers in solving complaint-related problems, but they are highly mobile and work at least twice a week away from desktop computers. Engineers, on the other hand, are involved in routine work on the mill premises, which is characterized as micro mobility. These people are most often overloaded with customer's questions, are heavily involved in the communication process and need a good visual interface when describing laboratory results to the customers. Effective customer response should be immediate. No matter how mobile and where the customers are, they expect to have problem solutions on-line (PC and/or mobile device) and in real time. Wireless has value for managers on the move, whereas high-resolution desktop computers are mandatory for engineers.

Temporality. Systems such as electronic mail or mobile phone calls may have several advantages; they are likely to be a burden in the complex group work activity of handling complaints. The above systems are very intrusive by their very nature, especially in communication-intensive environments when people have to work under the pressure of time. E-mail is an asynchronous medium, but imposes information overload with risk of being unable to keep track of how the handling of the complaint is progressing. Another problem is related to telephone calls. While it is an extremely flexible medium, producing lots of unstructured and codified information, it also requires the user to do a lot of work by manually retransmitting that semi-structured information to the relevant team member. Knowledge workers in the paper-manufacturing company spent 15-30% of their day searching for information and lose about 6 working hours altogether per complaint because of interruptions to their work. People often became frustrated because they were unable to get to things via a single access point, as necessary information was often spread across different actors and communicated via multiple devices. Electronic media may lead to new patterns of communication, encouraging people to collaborate at the point of need with certain team members, perceiving each other's actions instantly but asynchronously.

Fragmentation of information. In our study the notion of information fragmentation refers to what and how activity relevant information is communicated, shared and stored by the group members. Variety in the communication medium used makes the information flow process slow and inefficient (knowledge workers spend a lot of time on requesting and providing information manually, a lot of copy/paste work is involved). The system fails to integrate and structure the

incoming information from the customers and leaves it optional for the customers to enter. This also makes it more difficult to perform statistical analysis of annual complaints and identify the various problems associated with it.

We mentioned earlier that collaboration implies that the object of work should be common to the actors, i.e. by having access to a common artifact. In our case this means that information should be shared. This raises the question of what type of information and how it should be shared in order to support effective communication and shared understanding between actors. We believe that structured information, visible flow and sharing constitute a first underlying requirement here.

4. SYSTEM TO SUPPORT ACTIVITY-CENTERED COLLABORATION IN COMPLAINT HANDLING

4.1 Activity-Centered Collaboration

Malone [14] suggests that, in general, organizational technologies consist of production and coordination technologies that have implications for different processes. Complaints are regarded as service messages and clearly form part of management processes; they can therefore be enhanced by improved availability and communication of information. Work redesign can be based on the "object-oriented" modeling techniques discussed by Crowston et al. [5], where information flow can be streamlined by new technologies to provide necessary coordination among group members. In their recent study, Geyer and colleagues [10] introduced the notion of "object-centric sharing", where users collaborate in a lightweight manner, by aggregating and organizing different shared artifacts into larger collaborative activities with dynamic membership, hierarchical object relationships, and enable both real time and asynchronous communication. The "shared objects" paradigm of collaboration-aware content is very similar to Dourish's [8] notion of "placeless" documents, which define a set of people who are allowed to access the information, indicate who is currently looking at the content and allow broadcasting of data to members of the object. In our case, different actors involved in resolving the same complaint could use the data stored in one centralized data base and share a common virtual work environment accessible via different interfaces, rather than each having partial information or passing information among themselves. It is also argued that IT plays an important role in recent attempts to focus on core activities of reengineered business processes. The organization should be less centered on the process actors, and more on the activities they have to perform. The emergence of mobile communication technologies provides the relevant support for flexible forms of organization and temporal group work in producing a service. Systems for activity-centered collaboration which could also incorporate mobile technologies would appear to correspond and facilitate new forms of

customer service, offering the kind of support for all geographically dispersed group members.

4.2 Web-based System with Mobile Access

Nowadays the term IP technology convergence is quite popularly used. By this is meant a certain kind of fusion or close interaction of the Internet and mobile phone networks. This wireless access to the digital content of the Internet via mobile devices, including mobile phones, smart phones, PDA's, handheld computers (communicators) and portable PCs (laptops) is understood as the *mobile Internet*. From the convergence of Internet-based systems and mobile computing technologies, the term mobile Internet systems was derived, which is sometimes replaced by mobile web-based applications or M-commerce. Let us describe the architecture of the proposed complaints handling system with mobile functionality (see Figure 1). A web-based system is primarily built for a PC and GPRS-enabled smart phone, but is written with generic code and will in fact work on all html\xhtml-supporting devices.

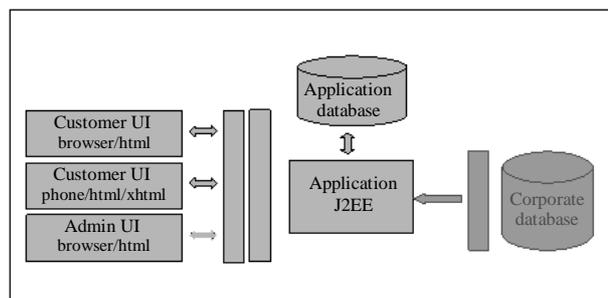


Figure 1. Architecture of the proposed system

Customer/Admin UI. The customer/administrator uses a web browser on a computer or a phone to access our system. The browser of a mobile device does not have to support xhtml, but it must support html. It means that we use html (html4-strict) instead of xhtml both for the PC and the phones. Since xhtml is easier to implement than html, it is usually the case that a smart phone browser that supports html also supports xhtml.

Servlet/interface. The interface between the browser and our system is based on servlet middleware or Java servlets and uses a template system, named "Velocity", to create the actual interface.

J2EE. The servlets use an abstraction layer that we have created and named "Knowledge Base", which accesses the J2EE-based core of the system. This is the application core, which is where all business logic is handled.

Application database. The application core stores all data in a database. All access to the database goes through an abstraction layer, which makes it possible to change the database to another brand without having to recompile the application.

The web-based complaints handling system can be accessed via a mobile device and the Internet accompanies its user everywhere, regardless of the type of device used. The system can work on any html\xhtml-supporting mobile device, but the Nokia 6600 was chosen for its reliability, faster navigation and cost efficiency compared with alternative smart phones. In a final product the web-interface and mobility feature should be integrated on top of the existing corporate ERP system.

4.3 New Communication Structures and Implications

The main users of the application will be key customers (resellers, purchasers, printing managers, plant manager) and paper-mill employees involved in the complaints process at the mills (engineer and manager of Technical Customer Service Center (TCS), director of TCS, logistics manager, production and operations managers, mill management and sales office personnel.

The main features of the proposed complaints handling system are traceability and structured information. Traceability will allow the user of the application to track the progress of complaints handling activities in real time, the ability to access and update the complaint file – a self-documenting and self-reporting common most important document within a system – via a PC or any other mobile device. When somebody is working on something with the complaint, the customer as well as other colleagues can see it. The principal value for the customer will be the transparency of the handling process, which is critical to them - the ability to track the complaint in real time and to have control over the process from the date the complaint was instigated.

Show all complaints
Show discussions (1)

Progress of complaint #000020004259 , summary

Time	Task	Status
16.6. 22:00	Complaint	Filed
16.6. 22:00	Samples	Received
		16.6. 22:17
16.6. 22:22	Technical statement	Made
16.6. 22:22	Studies	Completed
-	Final decision [Make]	Pending
-	Response	Pending

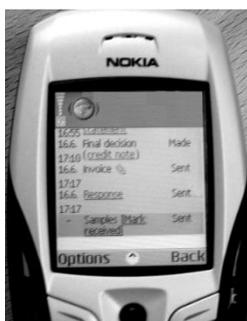


Figure 2. Traceability feature and system's visual outlook on PC and smart phone

Another important advantage of the web-based system is to have very structured initial information coming from customers. This also makes them think about the reasons before claiming and include additional info at the very beginning. A customer in an ideal case will create a complaint file at the very beginning of problem recognition, and should provide as much information as possible about the problem. The user will have the option of choosing a complaint type and filling in the required information fields, according to the reason for the complaint. In addition, we aim to ease the job of TCS at

the mill too. The system incorporates instantaneity and supports automatic notification; thus participants perceive each other's actions with no perceptible delay. Each group member is instantly notified by the system when a certain activity is finished or new questions arise. Depending on the work mobility, user may select whether he or she wants to receive e-mail notification or an SMS.

Usage scenarios. There are two scenarios for the modernized complaint system. In the best-scenario case, we have a direct link between a customer who files the complaint and the TCS (manager & engineer) at the mill. The second scenario is more applicable abroad only if language constraints do not allow the customer to send a complaint directly to the mill. In this case a sales office should register the complaint on the web during a phone call with the customer. A password will be required to log into the web-based system, so paper-mill employees working with complaints will know who the customer is. A self-documenting and self-reporting complaint file might include auto alerts in the form of an SMS or e-mail for actors performing at the end of the process, e.g. operations and productions managers, logistics managers, etc. At the end of the complaints handling process, the auto summary of the complaint file will generate a PDF file, and will allow the use of the well-structured information for various statistical analyses (beyond the current way of monitoring the success rate of complaint reply and closing time).

Referring to Figure 2, it is important to note, that the "visual outlook" on the smart-phones will be simplified, as usability of such mobile phones is limited and requires much more relevant and precise information than on a PC. We believe a smart phone serves merely as an additional tool and is not a replacement for a PC. It is meant to be used for information tracking, other checkups and decisions when a user cannot access the PC. Having the smart phone on hand, a user would simply be able to use the system when a PC is not an option. A PC/Laptop is also supposed to be the primary tool for customers filing complaints. However, they will also have the choice of making the best use of smart phones when checking the status of very urgent complaints.

5. CONCLUSIONS AND FURTHER RESEARCH

We have entered the field of research into what business currently needs and what consulting companies do not take seriously. In this paper we have described a mobile web-based collaborative system that we have designed and developed for managing complaints in the collaborative group-work environment of the fine-paper industry. The benefits - including stronger customer relationships, lower operating costs and better use of human resources - are available to any complex customer service activity that uses collaborative group work and seizes the initiative via an activity-centered collaborative system with mobile access.

First of all, collaborative customer service reduces costs while solving problems related to spatiality, temporality of working time and fragmentation of information. The math is simple. Just compare the cost of receiving and responding to several hundred phone calls and e-mails to the cost of sending a single, well-crafted message to a targeted audience. Secondly, a collaborative and always accessible system allows TCS to support more customers and resolve more complaints with existing staffing levels. With automated proactive service mechanisms carrying more of the burden, TCS no longer has to grow in proportion to demand. A well-structured and transparent way for handling routine issues also frees TSC staff to focus on more complex and/or urgent problems that truly warrant their personal attention and knowledge workers' skills. As a result, instant and transparent way of handling critical incidents is a key to a long-term customer relationship based on trust, customer empowerment and smooth service experience.

Despite the obvious positive effect of using an activity-centered collaborative system for handling complaints, various barriers do exist to its widespread adoption and use. The client of a new system is the worker at customer service and only indirectly the organization. A successful design is one that increases the group members' satisfaction and improves their attitudes towards work. The major barrier to introducing a state-of-the-art technology in an organization is the learning effort required: this is why we need "usable" systems and devices. On the other hand, people are likely to learn how to use new technology if they find it useful. Our future studies will focus on evaluating the usability of the working prototype in field experiments, depending on the type of devices adopted by different team members. This form of evaluation is needed to develop the system further and to make sure it is usable and easy to use.

REFERENCES

- [1] Bellotti, V. & Bly, S. (1996) Walking away from the desktop computer: distributed collaboration and mobility in a product design team. In CSCW 96 proc. of the Conference on Computer Supported Cooperative Work, September 17-20, Boston, Mass.:ACM Press. p. 209-218.
- [2] Bertalanffy, L.V.(1962). General System Theory – A Critical Review. General Systems 7, 1-20.
- [3] Bui, T.X., Ondrus, J., "M-Computing for Real-Time Negotiation Support", in *Proc. of the 35th Hawaii International Conference on System Sciences*, January 2003.
- [4] Cho, Y., Im, I., Hiltz, R., and Ferjemstad, J., "An Analysis of Online Customer Complaints: Implications for Web Complaint Management," *Proc. of the 35th Hawaii International Conference on System Sciences*, January 2002.
- [5] Crowston, K., Malone, T. W., and Lin, F. (1987). Cognitive science and organizational design: A case study of computer conferencing. *Human Computer Interaction*, 3, 59-85.
- [6] Churchill, E. F. and Wakeford, N. (2001) Framing Mobile Collaboration and Mobile Technologies. In Brown, B., Green, N., and Harper, R. (Eds) *Wireless world: social and interactional implications of wireless technology*. Springer Verlag.
- [7] Divitini, M., Farshchian, B.A., Samset, H., "UbiCollab: collaboration support for mobile users", *Proceedings of the 2004 ACM symposium on Applied computing*, Nicosia, Cyprus
- [8] Dourish, P., Edwards, W. K., LaMarca, A., Salisbury, M., "Presto: an experimental architecture for fluid interactive document spaces", *ACM Transactions on Computer-Human Interaction (TOCHI)*, v.6 n.2, p.133-161, June 1999
- [9] Fussell, S. R., Kraut, R.E., Lerch, F.J., Scherlis, W.L., McNally, M.M., Cadiz, J.J., "Coordination, Overload And Team Performance: Effects Of Team Communication Strategies", *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, Seattle, Washington, United States, pp. 275 - 284
- [10] Geyer, W., Vogel, J., Cheng, L. T., Muller, M., "Supporting Activity-centric Collaboration through Peer-to-Peer Shared Objects". In *the Proc. of ACM SIGGROUP*, pp. 115--124, Sanibel Island, FL, USA, November 2003.
- [11] Lamming, M., Eldridge, M., Flynn, M., Jones, C., & Pendlebury, D. (2000) Satchel: Providing access to any document, any time, anywhere. *ACM Transactions on Computer-Human Interaction*, 7(3), Special Issues entitled "Beyond the Workstation: Human Interaction with mobile Systems", p. 322-352.
- [12] Litiu, R., Zeitoun, A., "Infrastructure for Mobile Collaboration ", in *Proc. of the 3rd Hawaii International Conference on System Sciences*, January 2004.
- [13] Luff, P., Heath, C. (1998) "Mobility in Collaboration", in *Proceedings of CSCW '98: Conference on Computer-Supported Cooperative Work*. ACM Press, 305-314.
- [14] Malone, T. W., "Modelling Coordination in Organisations and Markets," *Management Science* (33: 10) October, pp. 1317-1332.
- [15] Perry, M., O'Hara, K., Sellen, A. Harper, R. & Brown, B.A.T. "Dealing with mobility: understanding access anytime, anywhere." *ACM Transactions on Computer-Human Interaction (ToCHI)*, 4(8), 2001, pp. 1-25.
- [16] Paavilainen, J., "Mobile Business Strategies: Understanding the Technologies and Opportunities", 1st Edition, Addison-Wesley, GB, 2002.
- [17] Reichheld, F. F. and Schefer, P., "E-Loyalty: Your Secret Weapon on the Web," *Harvard Business Review*, July-August 2000, pp. 105-113.
- [18] Sterne, J., "Customer Service on the Internet: Customer Support, Direct Marketing, and Up-Selling". 2nd Edition, John Wiley & Sons, Inc., New York, 2000.

MOBICORS-Movie: A MOBILE COntents Recommender System for Movie

Jaekyeong Kim¹, Yoonho Cho², Seungtae Kim¹

¹ School of Business Administration, Kyung Hee University, Seoul, 130-701, Korea

² School of e-Business, Kookmin University, Seoul, 136-702, Korea

jaek@khu.ac.kr, www4u@kookmin.ac.kr

ABSTRACT

In spite of the rapid growth of mobile multimedia contents market, most of the customers experience inconvenience, lengthy search processes and frustration in searching for the specific multimedia contents they want. These difficulties are attributable to the current mobile Internet service method based on inefficient sequential search. To overcome these difficulties, this paper proposes a MOBILE COntents Recommender System for Movie (MOBICORS-Movie), which is designed to reduce customers' search efforts in finding desired movies on the mobile Internet. MOBICORS-Movie consists of three agents: CF (Collaborative Filtering), CBIR (Content-Based Information Retrieval) and RF (Relevance Feedback). These agents collaborate each other to support a customer in finding a desired movie by generating personalized recommendations of movies. To verify the performance of MOBICORS-Movie, the simulation-based experiments were conducted. The experiment results show that MOBICORS-Movie significantly reduces the customer's search effort and can be a realistic solution for movie recommendation in the mobile Internet environment.

Keywords: mobile commerce, recommender, collaborative filtering, agent, relevance feedback

1. INTRODUCTION

As mobile Internet technology becomes more increasingly applicable, the mobile multimedia contents market has recorded remarkable growth. In spite of this rapid growth, however, most of the customers experience inconvenience, lengthy search processes and frustration in searching for the specific multimedia contents they want [3]. These difficulties are attributable to the current mobile Internet service method based on inefficient sequential search:

1 When customers log on to the mobile Internet site using a mobile phone, they are presented with the best-seller list or a list of multimedia contents in sequence of most recent introduction.

1 Customer pages through the list and selects an entry to check out its contents hoping that the contents will be what he or she wants.

1 If the customer likes the contents, he or she may make a purchase. Otherwise, the customer repeats the same steps until the customer either has stumbled over the right one or decides to give up.

Using current service method, the expected number of contents the customer views before he or she hits the desired movie far exceeds the acceptable level. Inconvenience, lengthy search processes, and frustration make customers easily annoyed by the amount of irrelevant and uninteresting information they have to filter out. During a search, as more time is consumed, frustration builds with the result that the purchase conversion rate drops. To make search processes more acceptable, a more efficient searching aid that suggests only the movies meeting the individual customer's preference is needed. This paper proposes a MOBILE COntents Recommender System for Movie

(MOBICORS-Movie), which is designed to reduce customers' search efforts in finding desired movies on the mobile Internet. The system combines two of the most popular information filtering techniques: Collaborative Filtering (CF) and Content-Based Information Retrieval (CBIR).

MOBICORS-Movie consists of three agents: CF, CBIR and RF (Relevance Feedback). These agents collaborate each other to support a customer in finding a desired movie by generating personalized recommendations of movies.

The simulation-based experiments were conducted to verify the performance of MOBICORS-Movie. The experiment results show that MOBICORS-Movie significantly reduces the customer's search effort, so it can be a realistic solution for multimedia contents recommendation in the mobile Internet environment. We believe that deployment of MOBICORS-Movie offers the following benefits to both consumers and suppliers of mobile multimedia contents:

1 Customers can purchase mobile contents with much less search effort and much lower connection time to the mobile Internet, because they can much more easily find desired mobile contents.

1 Mobile contents providers can improve the profitability of their business because lower customer frustration in finding desired contents increases revenue through an improved purchase conversion rate.

2. EXISTING TECHNIQUES

2.1 Collaborative Filtering

The recommender system is one of possible solutions to

searching for individually preferred contents from a large-contents database. A recommender system is defined as a system that assists customers in finding the items they would like to purchase. One of the most successful recommendation techniques is Collaborative Filtering (CF) [7,8], which has been widely used in a number of different applications.

Collaborative filtering is an information filtering technique that depends on human beings' evaluations of items. It is an attempt to automate the "word of mouth" recommendations that we receive on a daily basis from our family, friends, and colleagues. It identifies customers whose tastes are similar to those of a given customer and it recommends items those customers have liked in the past. In general, CF-based recommender systems make recommendations according to the following steps [4,7]: (1) A customer provides the system with preference ratings on items that may be used to build a customer profile. (2) The system applies statistical or machine learning techniques to find a set of customers, known as neighbors, who had in the past exhibited similar behaviors (i.e., either they had purchased a similar set of items or they had given the set similar ratings). A neighborhood is formed based on the degree of similarity between a target customer and other customers. (3) Once a neighborhood is formed for a target customer, the system generates a set of items that the target customer is most likely to purchase by analyzing the items in which neighbors have shown an interest (*top-n recommendation*).

Although the CF is the most successful recommendation technique, it suffers from two major shortcomings. First, when there is a shortage of ratings, CF suffers from a sparsity problem [1,2,4,7]. Most similarity measures used in CF work properly only when there exists an acceptable level of ratings across customers in common. An increase in the number of customers and items worsens the sparsity problem, because the likelihood of different customers rating common items decreases. Such sparsity in ratings makes the formation of neighborhoods inaccurate, thereby resulting in poor recommendations. Next, CF suffers from a new item problem [1,5]. Since CF recommends an item based on customers' ratings on the item, it does not recommend a newly introduced item until some ratings of the item become available. The new item problem becomes even worse when the turnover rate of items is high. The Third issue is related to *scalability* [4,7]. Recommender systems for mobile customers have to deal with large amount of multimedia contents and millions of customers. Because these systems usually handle very high-dimensional profiles to form the neighborhood, the nearest neighbor algorithm is often very time-consuming and scales poorly in practice.

2.2 Contents-based Information Retrieval

Content-based information retrieval represents a movie

as a point in the multi-dimensional feature space and performs similarity-based retrieval using its contents-based keywords. In CBIR, customer describes movies using a query that is a set of example keywords. A query is internally represented as multiple points (i.e. query points) that have keywords of example movies. In general, CBIR systems retrieve movies according to the following steps [3,9]: (1) A system presents a query to the customer via the results of collaborative filtering as a request for desired movies. (2) The system searches for movies similar to the query. The similarity between a movie in the database and a query is calculated using the distance between corresponding points in the feature space. (3) The movies with the highest degree of similarity are retrieved and recommended to the customer.

In spite of the virtues of CBIR in retrieving movies similar to a query, CBIR rarely brings a customer to the desired movies immediately. The reason for this is that any combination of example movies may not precisely represent the movies that a customer desires. For a system to handle this gap properly, it needs the ability to learn about what movie the customer really wants through iterative interactions. The customer's current preference on the presented movies needs to be fed back so that CBIR can learn from this preference to retrieve, in the next iteration, movies more similar to the one customer really wants. This learning process, the *relevance feedback*, is an essential mechanism for a faster search of desired movies. The degree of preference of a movie in MOBICORS-Movie is expressed in three-level (e.g. *preferred, neutral, unpreferred*) weights. We will refer to a set of preferred movies as a *preferred set*. The movies in the preferred set are used for query refinement for the purpose of learning customer's current preference. And relevance feedback is performed using Rocchio's algorithm [6].

3. MOBICORS-Movie

3.1 Recommendation Procedure

MOBICORS-Movie is designed to reduce customers' search efforts in finding desired movies on the mobile Internet. The system consists of three agents: CF, CBIR and RF (Relevance Feedback). These agents collaborate each other to support a customer in finding a desired movie by generating personalized recommendations of movies. Figure 1 shows the recommendation procedure of MOBICORS-Movie.

First, an initial recommended movie list is generated by CF agent and shown up to the mobile customer. A customer selects one of the recommended movies using just a movie title. Then the customer sees the detailed information of selected movie, such as brief introduction of the movie, director, actors, actress, and so on. If he/she buys the movie, then he/she can watch the movie by mobile phone after payment process. If the customer

replied as “unpreferred”, the recommended movie list generated by CF agent except the selected movie is shown up to the mobile customer.

If the customer doesn’t buy the movie but replied as “preferred”, CBIR agent and RF agent search for similar

movie(s). Then the customer sees the detailed information of selected movie. The procedure is terminated until the customer buys the movie or disconnects the searching process.

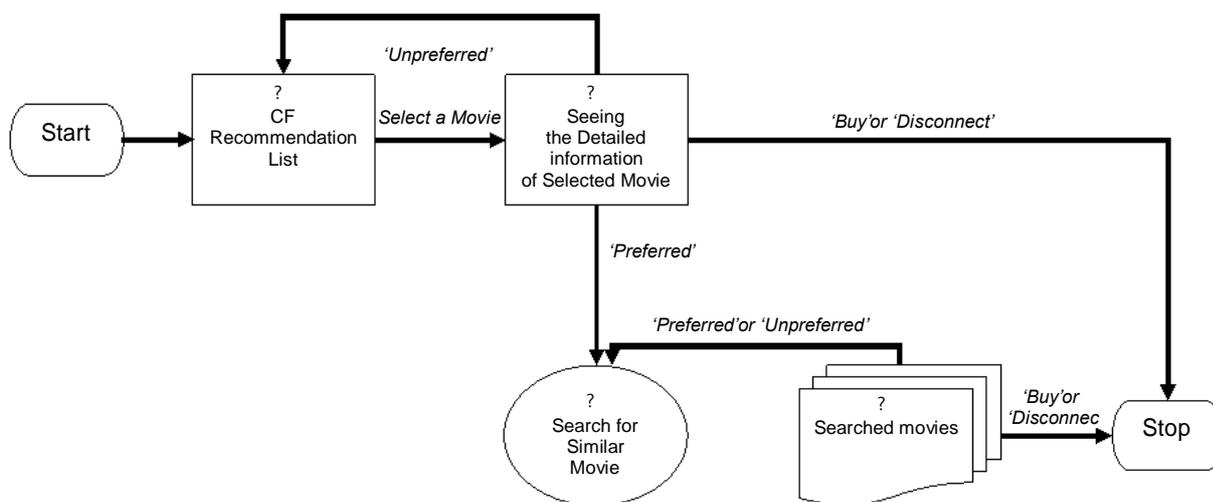


Figure 1. The procedure of MOBICORS-Movie

3.2 System Architecture

The architecture of MOBICORS-Movie is shown at Figure 2.

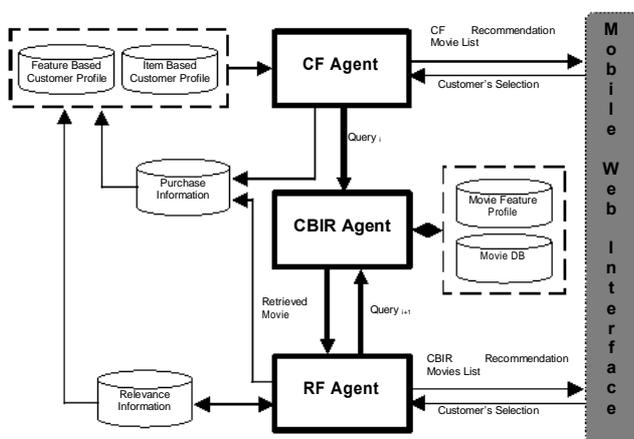


Figure 2. The architecture of MOBICORS-Movie

The CF agent generates a list of recommended movies and provides an initial movie to the CBIR agent. This agent creates the customer profile using purchase and preference information to identify neighbors and generate recommendations. When the CF-generated recommendation list is presented, a customer skims through the list to see if there are any movies of interest. Then, the customer selects an entry to view the movie.

After viewing, the customer may decide to purchase the movie or decide whether to use the movie as a query for content-based search of similar movies or to go back to the CF-generated recommendation list.

When the customer decides to use the viewed movie as a starting query for further search, the viewed movie is passed to the CBIR agent as an initial query, and the agent retrieves movies based on similarity between the query and other movies in the database. For all movies in the database, this agent calculates the similarities from the query and generates a list of the most similar movies as recommendations. It then passes the retrieved movies to the RF agent. RF agent presents the retrieved movies to the customer one by one, and interactively elicits the user’s preference judgment on the presented movies.

At any point in this presentation session, the customer may decide to buy a movie or decide to quit. After all of preference judgment, either preferred or unpreferred, are made, the RF agent updates the preference information and purchase databases with all the fed back preference and/or purchase information respectively for later use by the CF agent when the customer revisits the site. If all of presented movies are marked as unpreferred, the search session returns to the CF-generated recommendation list. Otherwise, the RF agent learns the customer’s current preference using his/her preference judgment, formulates the new query, and passes the query to the CBIR agent for the next iteration of retrieval.

3.3 Profile Management

Movie Profile: A movie profile includes information about the characteristics of movies. A movie profile is represented by matrix $V=(v_{jk})$, where j and k implies j th movie and k th keyword. If j th movie contains k th keyword, the value of v_k is 1, otherwise it is 0. For example, the number of movies is 7, and that of keywords is 6, an example of movie profiles is represented as Table 1.

Table 1. A Movie Profile

	Betrayal	Crime	Cult-favorite	Drama	Police	Twist-in-the-end
Godfather	1	1	0	1	1	1
Pulp Fiction	0	1	1	1	0	0
Reservoir Dogs	0	1	1	0	1	1
Fight Club	0	1	1	1	0	1
Fargo	0	1	1	0	1	0
Psycho	0	0	1	0	1	1
Goodfellas	1	1	0	1	1	0

Feature-based Customer Profile: A feature-based customer profile represents customer's preference on keywords. It is composed of customer-keyword matrix $F = (f_{ik})$, where f_k implies the preference degree of i th customer on k th keyword. For example, the number of customers is 4, and that of keywords is 6, an example of feature-based customer profiles is represented as Table 2.

Table 2. A Feature-based Customer Profile

	Betrayal	Crime	Cult-favorite	Drama	Police	Twist-in-the-end
Kim	5	10	0	5	-10	0
Cho	10	5	-5	5	5	10
Kang	5	15	5	10	-5	0
Lee	0	10	0	10	0	-5

Item-based Customer Profile: An item-based customer profile represents customer's preference on movies. It is composed of customer-movie matrix $P = (p_{ij})$, where p_{ij} implies the preference degree of i th customer on j th movie. For example, the number of customers is 4, and that of movies is 7, an example of item-based customer profile is represented as Table 3..

An item-based customer profile includes information about customer's preferences on movies. Generation of CF recommendations depends totally on item-based customer profile. MOBICORS-Movie builds the item-based customer profile using purchase records and fed-back binary preference information, either preferred or unpreferred.

Table 3. An Item-based Customer Profile

	Godfather	Pulp Fiction	Reservoir Dogs	Fight Club	Fargo	Psycho	Goodfellas
Kim	5	15	0	15	0	-10	5

Cho	15	-5	5	5	-5	0	10
Kang	15	10	15	15	15	0	5
Lee	10	5	5	10	5	-5	5

The profile is updated as the following equation:

$$p_{ij} = \begin{cases} +2 & \text{if customer } i \text{ has purchased movie } j \\ +1 & \text{if customer } i \text{ has marked movie } j \text{ as preferred} \\ -1 & \text{if customer } i \text{ has marked movie } j \text{ as unpreferred} \\ 0 & \text{if customer } i \text{ has not seen movie } j \end{cases}$$

Note that the brief introduction of movie that customer viewed from CF recommendation list and decided to use or not to use as an initial query for CBIR is counted as a preferred or unpreferred movie, respectively. As shown in above equation, the cells of the rating matrix have four possible different values. We place the highest ratings on the previously purchased movies, because they should reflect the customer's taste the most strongly. The movies marked as preferred or unpreferred are represented in ratings of half the magnitude of the purchased ones with an opposite sign. Assigning -1 to unpreferred images gives a better chance to unseen movies over unpreferred ones in generating recommendations. Please note that further research work is necessary to justify heuristic choice of values used for preference ratings, since their relative sizes in magnitude could affect recommendations of CF.

The ratings in customer profile are constantly replaced with newly obtained purchase and preference information to dynamically reflect customer's most recent preference. This is significantly different from the customer profiles used in traditional CF techniques [7,8].

4. EXPERIMENTAL EVALUATION

4.1 Experiment

For the purpose of the performance evaluation of MOBICORS-Movie, we developed a Web-based application system running on a PC with exactly the same user interfaces as the mobile phone-based MOBICORS-Movie system. Using the system, we carried out the experiments with the intent of answering three major questions:

- (1) How much performance improvement does MOBICORS-Movie deliver as his/her search and purchase information in customer profile is accumulated?
- (2) How does the number of movies CBIR agent search affect the performance of MOBICORS-Movie?
- (3) How much performance improvement does MOBICORS-Movie deliver compared to other recommender systems in mobile phone environment?

For the experiments, the 250 movies from IMDB(www.imdb.com) were used. MOBICORS-Movie used the top 100 keywords sorted by frequency from

IMDB. Virtually generated 1000 mobile Web customers are participated in the experiment. They are set to purchase 5 movies randomly, and their initial purchase information is stored as an initial feature based customer profile and item based customer profile.

For performance evaluation of MOBICORS-Movie, a metric *views-per-success* (*vps*), defined as the number of movies viewed by a customer before he/she purchases a movie in a search session, was devised. The *vps* measures the amount of effort a customer takes per successful search [3].

4.2 Results and discussion

The quality of recommendations of CF is known to vary by the size of the neighborhood [4,7]. The neighborhood size yielding optimal performance depends on customer profile, whose contents change all the time. For this reason, comparison between optimal performances of different periods requires increasing number of additional experiments to calculate the optimal neighborhood for every period. However, considering the burden of participants in the experiment, we used the optimal neighborhood size of the first period and fixed the neighborhood size for the rest of the experiment. According to Figure 4, the optimal neighborhood size was determined to be 40. For the rest of the experiment, we fixed the neighborhood size at 40.

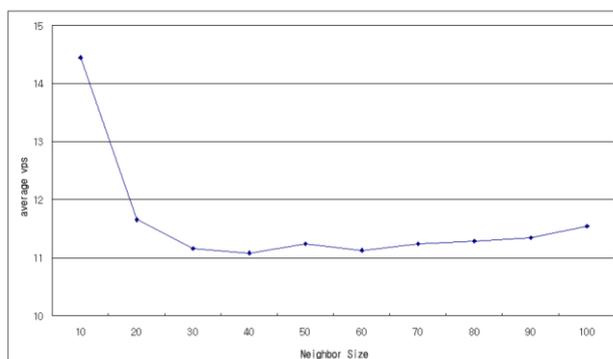


Figure 3. Average *vps* over diverse neighborhood sizes

To answer the question 1 and 2, we observed the variation in *vps* over periods and n , number of nearest neighbor movies retrieved in CBIR. This is observed as a decreasing curve in Figure 4, which shows the overall performance of MOBICORS-Movie in terms of *vps* vs. period for five different values of n .

As periods move on, more rating information becomes available and the customer profile with more ratings makes neighborhood formation more accurate, thereby improving the quality of CF recommendations. Figure 4 indicates that the variation in *vps* by n is significant. Furthermore, it is observed as the difference in *vps* by values of n especially in early periods, and as the diminishing trend of difference in *vps* over periods in

Figure 4.

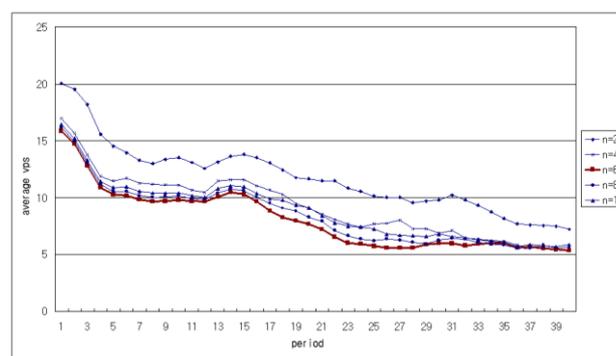


Figure 4. Overall performance of MOBICORS-Movie

Furthermore, as the value of n becomes from 2 to 6, the performance becomes better, but the value of n goes over 6, the performance becomes worse. In general, sparsity level of customer profile and heterogeneity in customers' buying behaviors are the most important factors influencing the effectiveness of CF [2,3,4]. In cases with high level of either sparsity or heterogeneity, CF's performance is poor, thus we select optimal value of n . The value of n can be selected considering other factors such as customer's burden of feedback in CBIR or the firm's marketing strategy toward the diversity in recommended movies.

To answer the question 3, the performance of MOBICORS-Movie is compared to those of two other recommender systems, the best-seller-based system (Bestseller) and a typical CF-based recommender system (pure-CF), whose procedure is identical to that of CF-based recommendation of MOBICORS-Movie except that, in building a customer profile, it uses purchase information only.

As compared in Figure 5, the *vps* of MOBICORS-Movie is about 60% and 72% lower, at 40 periods, than that of the pure-CF and best-seller-based systems respectively. In other periods, the performance gap between MOBICORS-Movie and other techniques is higher. Figure 5 also shows that the rates of improvement in *vps* over the 10 periods (i.e. learning speed of the system) of MOBICORS-Movie, pure-CF and best-seller-based systems are 50%, 68% and 55% respectively. These results show that, of the three systems evaluated, MOBICORS-Movie offers not only the lowest *vps* but also the most robust performance over 10 periods. The superior performance of MOBICORS-Movie over the pure-CF results from MOBICORS-Movie's accelerated learning of customer preference from additional preference rating information fed back from CBIR. This evidences that MOBICORS-Movie, as intended, successfully overcomes the sparsity problem of CF.

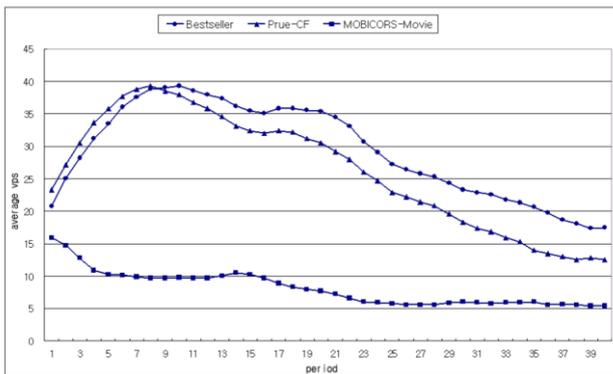


Figure 5. Comparison of MOBICORS-Movie and Other Recommendation Techniques

5. CONCLUSION

It is devised a recommender system, MOBICORS-Movie, to address the real problems encountered in movie watching services in the mobile Web environment. The system combines two techniques from different research domains: recommender system and information retrieval. By employing two techniques, MOBICORS-Movie generates recommendations with the following characteristics: (1) MOBICORS-Movie recommends movies by reflecting the opinions of other customers with similar tastes. (2) MOBICORS-Movie recommends movies with similar keyword features to the ones the customer currently prefers. (3) MOBICORS-Movie recommends movies by learning the customer's preference adaptively via feedback on recommended movies.

To verify the performance of MOBICORS-Movie, the simulation experiments were conducted. From the experiment result, we can conclude that MOBICORS-Movie is a viable solution to the problems currently encountered in movie watching on the mobile Web, and that it can be expected to reduce the search effort, thereby increasing the purchase conversion rate. Use of MOBICORS-Movie offers the following benefits to both consumers and suppliers of mobile contents: (1) Customers can purchase contents with much less search effort and much lower connection time to the mobile Web, because they can much more easily find desired mobile contents. (2) Mobile contents providers can improve the profitability of their businesses because

lower customer frustration in finding desired contents increases revenue through an improved purchase conversion rate.

With the rapid growth of the mobile Web service, the mobile Web-based recommender system for other types of multimedia contents, such as music on demand (MOD) will continue to be an area of research interest in the future.

REFERENCES

- [1] Balabanovic, M., Y. Shoham, "Fab: Content-based Collaborative Recommendation", *Communication of the ACM*, Vol. 40, No. 3, pp 66-72, 1997.
- [2] Cho, Y.H., J.K. Kim, S.H. Kim, "A Personalized Recommender System Based on Web Usage Mining and Decision Tree Induction", *Expert Systems with Applications*, Vol. 23, No. 3, pp 329-342, 2002.
- [3] Kim, C.Y., J.K. Lee, Y.H. Cho, D.H. Kim, "VISCORS: a Visual Contents Recommender System on the Mobile Web", *IEEE Intelligent Systems, Special issue on Mining the Web for Actionable Knowledge*, Accepted for publication, 2004.
- [4] Kim, J.K., Y.H. Cho, "Using Web Usage Mining and SVD to improve E-Commerce Recommendation Quality", *LNAI 2931: Intelligent Agents and Multi-Agent Systems*, pp 86-97, 2003.
- [5] Melville, P., R.J. Mooney, R. Nagarajan, "Content-boosted collaborative filtering for improved recommendations", *Proc. the 18th National Conference on Artificial Intelligence*, pp187-192, 2002.
- [6] Rocchio, J.J., "Relevance Feedback in Information Retrieval", In Salton. G., editor, *The SMART Retrieval System - Experiments in Automatic Document Processing*, pp.313-323. Prentice Hall Inc., Englewood Cliffs, NJ., 1971.
- [7] Sarwar, B., G. Karypis, J.A. Konstan, J. Riedl, "Analysis of recommendation algorithms for e-commerce", *Proceedings of the ACM Conference on Electronic Commerce*, pp 158-167, 2000.
- [8] Schafer, J.B., J.A. Konstan, J. Riedl, "E-commerce recommendation applications", *Data Mining and Knowledge Discovery*, Vol. 5, No. 1-2, pp 115-153, 2001.
- [9] Zhou, X.S., T.S. Huang, "Relevance feedback for image retrieval: a comprehensive review," *ACM Multimedia Systems Journal*, Vol. 8, No. 6, pp 536-544, 2003.

Mobile Payment System with Privacy Protection

Kuenliang Sue, Wanpu Hsieh

Department of Information Management, National Central University, Chung-Li, 320, Taiwan, China
{klsue, im880703}@mgt.ncu.edu.tw

ABSTRACT

Trade security is the main point of the payment system. The system should fit the security issues such as confidentiality, authentication, and non-repudiation. However, the privacy problem seems to be ignored. This article proposes a new structure to improve the privacy of mobile payment system when consumers buy something via mobile device. The new system needs to fit three points: familiarity, privacy, and non-repudiation.

Keywords: mobile payment, non-repudiation, privacy, mobile commerce

1. INTRODUCTION

The trade behaviors are changing gradually because of the rise of mobile commerce. In the past trade activity, people needed to bring a lot of cash by themselves. Gradually, the applications of the plastic currency grow, people reduce the times using cash. Then, the appearance of e-commerce, the trade is no longer confined to face-to-face. The trade no longer needs to stay in one regular place. However, some problems still exist. Although plastic currency is convenient, people need to take a lot of cards with themselves. Even the trade of e-commerce can be remote, the security and convenience are still insufficient. Fortunately, mobile payment can change the predicament. People do not need to bring many cards because the tool of the trade is the mobile communication device that we can find easily on our person. Convenience of the trade increases greatly. The trade activity can be processed in any time, at any where, and with any people. Nevertheless, due to the congenital defect of the mobile device, it is unable to carry on the complicated security mechanism. The security problem is the key point in mobile payment system. Figure 1 shows the traditional payment model of credit card. This is the payment model of e-commerce, too. With this model, we can understand the trade activity in the past.

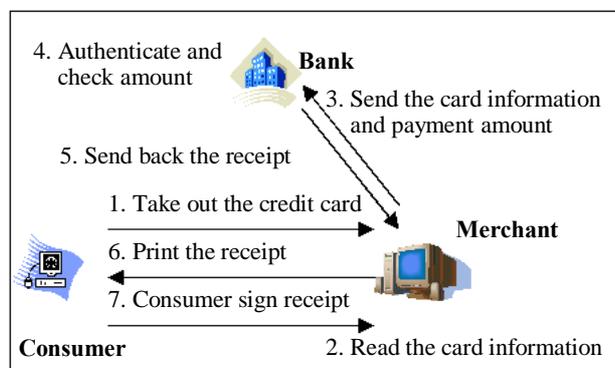


Figure 1. Traditional credit card payment model

Step 1, consumers take out the credit card to begin the

trade. Step 2, the merchant uses a card reader to read the credit card information of consumer. Step 3, merchant sends the credit card information to the bank. Step 4, the bank authenticates the consumer and checks amount to finish the payment. Step 5, the bank sends the receipt back to confirm the trade. Step 6, after the merchant receives the receipt, merchant will print out the trade receipt to do a confirmed by consumers. Step 7, consumers need to sign and confirm on the receipt. After finishing those steps, trade behavior by using credit card completes.

Although using plastic currency is very general for people nowadays, it seems that the whole procedures do not pay enough attention to the security issues. Potentially, one can pretend himself is someone else and forge a trade by using other people's credit card number. The similar situations also happen on the network. Except for worrying about the criminal gang to record the number of the card, consumers also need to believe the merchant very much and think that the merchant will not steal the personal information of the consumer. Because we must transfer the personal credit card to the merchant to help us read the information and finish the trade procedure. Although we can use receipts to confirm whether the amount of money of consumption and goods have mistakes or not, the important private information like trade account may be recoded and collected during the step of reading the card. Hence, the evil-minded merchant can use the information, forge into consumers and consumers are unable to be aware of it.

In order to remedy these known defects, this article proposes an improved architecture and meets the following demands. First, follow the past trade habits of the consumer, but change the paying device into the mobile community device or other hand-held community device. Secondly, strengthen privacy protection and disable the merchant to learn account information and accumulate shopping habits of the consumers. Also, the telecommunication operator (operator) is unable to know what the user bought. Last, offer the non-repudiation trade security.

2. RELATED WORKS

This section will review the related works of mobile payment system. In wireless mobile payment system the greatest challenge is the security threat. The study of Welch and Lathropped [1] divided these attacks into seven kinds as shown in Table 1. Except for unauthorized Access and Replay, others are set up on the basis of collecting and analyzing the package. Moreover, the Session High-Jacking attack could take the session instead of the user.

Table 1. The attack of wireless security

Traffic Analysis	Analyze the transmitting flow in the network
Passive Eavesdropping	Collect and analyze the network package to understand its content
Active Eavesdropping	Attract the destination to sending the specific package voluntarily to collect and succeed in analyzing packages
Unauthorized Access	The abnormal authorizes
Man-in-the-middle	Steal a glance or modifies the package that destination send
Session High-Jacking	Get involved and replace the session of the destination
Replay	Repeat sending the used package

Hence, wireless network also need to consider several important issues in the network security, such as confidentiality, authentication, authorization, integrity, non-repudiation. The best solution relies on encryption and the authentication mechanisms. Encryption mechanisms can offer the confidentiality and integrity. Authentication mechanisms can offer authentication, authorization, non-repudiation. In addition, the personal privacy of the consumer, e.g. shopping habits, is important security issues, too.

In order to apply mobile payment to the commercial trade, several trade systems considering about security issues are proposed. The research of Soliman and Omari [2] propose a dynamic encryption mechanism, the encryption key does not need regular storing in user's device, but produce the encryption key dynamically to avoid the encryption key lost or stolen. This system adopts symmetrical encryption to maintain confidentiality of consumer's trade and the encryption key exchange mechanism is also proposed to overcome the problem about losing encryption key. Later, Kungpisdan, Srinivasan, and Le propose an encryption key management mechanism [3] to improve the

symmetrical encryption problem that may be lost of key if the encryption key does not change. Maybe this kind of systems can protect confidentiality, but the importance of non-repudiation is not mentioned.

In the model of the trade, non-repudiation is a very important security mechanism. All of the symmetrical encryption systems are unable to offer non-repudiation. Thus, Fourati, Ayed, Kamoun, and Benzekri propose a research that hope to take SET [4] mechanism to provide non-repudiation and personal privacy [5]. After consider the operation ability of mobile device, take the authentication mechanism of the part of SET, and combine WTLS/TLS [6] [7] to build a non-repudiation system which mixes the advantage of the two. Except for adopting the method like this, there are also systems introducing the mechanism of digital signature to improve performance. Such as the research of Herzberg, it adopts the digital signature mechanism of DSA [8] to satisfy non-repudiation.

In addition, the privacy of consumer also needs protection in the trade. The research of Sue adopts user identity (UID) to replace the cell-phone number [9] when the trade message transmitted to the merchant. Because consumers' user identity is regular, merchants probably collect customer's private information, such as habit about consumption. If the UID is not always the same, consumer's privacy can be protected. Based on the similar concept, Rubin and Wright [10] propose the variation code method in the payment system of the wired network. It utilizes random codes to prevent consumers' credit card number from spreading in the network. Combine the UID [9] and variation code methods [10] for wireless applications, the merchant probably can not accumulate consumers' shopping habit.

However, this method needs to assistance of the operator. Hence the operator has an opportunity to accumulate consumers' shopping habit. The main reason of the problem lies in the structural design. The mobile payment model is often modified by e-commerce payment model. Only the trade equipment of the front-ends is changed. No matter which payment model is adopted, the account numbers of the trade should pass to the merchant. Consumers will lose the privacy, such as shopping habits.

3. A NOVEL MOBILE PAYMENT SYSTEM

This article proposes a novel model to reduce the security risk of the credit card payment and provide privacy protection in mobile payment. The proposed system will employ mobile equipment as the front-end traded device and the trade information will not be transmitted via any merchants. Hence, the system is able to protect the security of the trade account. The structure of the system is shown as Figure 2.

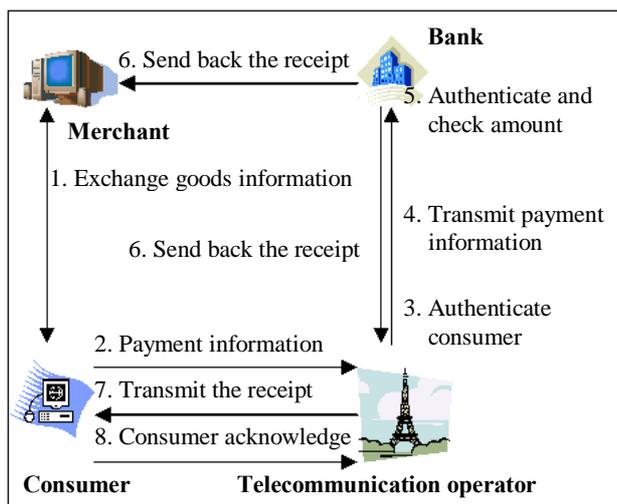


Figure 2. Structure of Payment System

3.1 Structure of Payment System

In order to exchange the trade message with consumer's mobile equipment directly, the trade structure demands some techniques to support, such as the Bluetooth interface. The interface is better an independent set to disable merchant from acquiring consumers' information, such as cell-phone number or account number. The banks in such system need to connect with the operators to deliver trade information to the users of the mobile payment system. In addition, consumers' mobile equipment needs the memory devices to store private key because this article applies consumers' private key to do digital signature to maintain non-repudiation. However, the signature algorithm is part of asymmetrical algorithms, several researches point out that the asymmetrical algorithms is more complicated than the symmetrical algorithms. Hence, we can select an asymmetrical algorithm to minimize the burden of mobile device when implementing the proposed payment system. According to the comparison [11] shown in table 2, the RSA algorithm is adopted to do digital signature.

Table 2. Digital signature perform comparison sheet[11]

Signature	Verifiable Encryption	
	#exponentiations	Size (bytes)
RSA	7.5	400
Gennaro et al	7.5	400
GramerShoup	8.7	544
GQ	10.5	544
Schnorr	8.3	388
ElGamal	8.5	388
DSA	11.6	484

In addition, we can adopt shorter private key on the mobile device to reduce the computation load and request the telecommunication operator to employ other

complicated digital signature algorithm for security after confirming users because the telecommunication operator has no restriction on computation. If consumers lose the mobile device, they just need to notify the operator stopping the service of the cell-phone. The forgers will be unable to pass the second layer authentication mechanism, and obtain the digital signature of the telecommunication operator. It can protect both the banks and consumers.

3.2 Procedure of Payment System

In this section, the trading procedures of the proposed system will be introduced. Table 3 defines the code to help readers understand.

Table 3. Code contrast

Name	Code
Merchant	Merchant
Consumer	Consumer
Telecommunication operator	Operator
Bank	Bank
Hash function	Hash
Item identity	Item_id
Item price	Item_price
Merchant identity	Merchant_id
Trade identity	Trade_id
Message Authentication Code	MAC
Cell-phone number	PhoneNum
Digital signature of Consumer	Sign_U
Digital signature of operator	Sign_O
Digital signature of the bank	Sign_B
Digital signature text of Consumer	Text_Usign
Digital signature text of Operator	Text_Osign
Digital signature text of the bank	Text_Bsign

The payment system has 8 steps. The motion of each step and content of each message will be introduced below

Step 1, (Item_id, Item_price, Merchant_id, Trade_id) + MAC

Consumers communicate with the trade system of the merchant and get the trade message from it via the bluetooth interference. In order to prevent the trade information from being modified, the merchant will use hash function to product message digest, such as MD5 [12] or SHA-1 [13]. The detailed composition of message is shown in Figure 3.

Hash

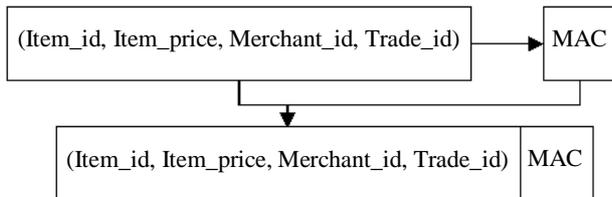


Figure 3. Trade procedure - Step 1

Step 2, $\text{Sign}_U(\text{Trade_id}, \text{MAC}) + \text{PhoneNum}$

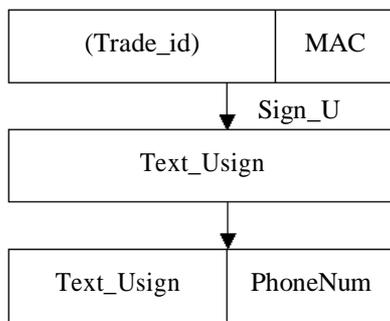


Figure 4. Trade procedure - Step 2

Consumers' mobile device will receive the trade message from the trade system of the merchant. Then consumer takes out the trade identity and MAC, and signs to respond the trade. Transmit the message to the telecommunication operator by GPRS or other wireless technology. The content of message is shown Figure 4. To prevent the operator and the bank from knowing the detailed content of the items that consumers bought, only the trade identity is transmitted.

Step 3, Authenticate consumer

When the operator receives the message that consumers transmitted, the operator will do the GSM authentication method to confirm the user identity of the mobile device.

Step 4, $\text{Sign}_O(\text{Sign}_U(\text{Trade_id}, \text{MAC}), \text{PhoneNum})$

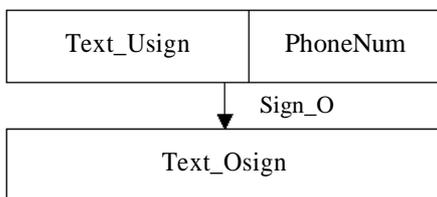


Figure 5. Trade procedure - Step 4

If consumers pass authentication, the operator will do digital signature to show the responsibility. In order to strengthen the non-repudiation, the operator can adopt more complicated digital signature algorithm. The content of the whole digital signature are shown in Figure 5. Later, transmit the message to the bank.

Step 5, Authentication and payment inside the bank

When the bank receives messages from the operator, the bank will verify operator and consumers by the digital signature of them respectively. Then, the bank confirms consumer's available amount and pays the money.

Step 6, $\text{Sign}_B(\text{Trade_id}, \text{MAC})$

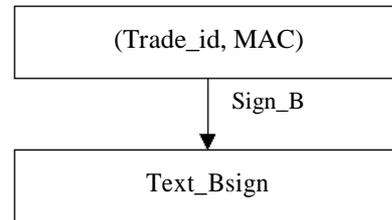


Figure 6. Trade procedure - Step 6

After finishing the payment, the bank produces the receipt of the trade and adds the digital signature to show the responsibility. The content of receipt is shown in Figure 6.

Step 7, $(\text{Text}_B\text{sign})$

This procedure only passes the receipt that is come from the bank. The operator does not need to add any digital signatures on the receipt because mobile device will need more complex computation if there are more digital signatures. In addition, the content of receipt is just trade identity, has no safety consider.

Step 8, consumers acknowledge the operator.

Step 8 is to ensure that consumers have already the receipt, but it is not the essential procedure. Such procedure can be omitted to simplify the procedure of the system and payment check can be done by the bank monthly.

Here showing some expected goals which this article tries to satisfy. First is familiarity. Although consumers have more one action, it can protect the individual privacy of consumer. Actually, the action is equivalent to the consumer signature in the past. Furthermore, such action can be simplified by software design, the consumer just need to press the next button when mobile device receives the trade message. Second is that the privacy of consumers is protected. The merchant only participates in Step 1 and Step 6 (Figure 2). These two procedures transmit the receipt and the trade message which include item identity, item price, and trade identity. The merchant are unable to know the account that consumers use for trade, and even can not accumulate shopping habit of consumers because consumers do not transmit the user identity or the cell-phone number to the merchant. As in the aspect of the operator and the bank, because consumers just transmit the trade identity in Step 2 (Figure 2), the operator and bank are unable to know what consumers buy and accumulate consumers' shopping habit. Hence, consumer's privacy of shopping habit and trade account

can be protected. Third is the non-repudiation. Three digital signatures of the consumer, the bank, and the operator are used in this system and represent non-repudiation of three participants. Consumers' digital signature makes consumers unable to deny the trade. Digital signature of the operator makes the operator can not deny ever passing the trade messages from authenticated consumers to the bank. Digital signature of the bank makes the bank can not deny approving the payment.

4. EVALUATION OF SYSTEM SECURITY

This section analyzes the system structure and system procedure to evaluate the system if fit the request of security. Evaluation will follow three aspects: bank, merchant, and consumer.

4.1 Confidentiality

Confidentiality is usually done by encryption, but this system does not apply the encryption algorithm. Here is another opinion of the confidentiality in trade. Because the message which is transmitted by consumers in the network has important information inside, such as bank account of consumers, trade items information. The confidentiality protects the message not to be watched. If the information does not be transmitted in the network or the message transmitted in the network does not include such important information. Then, encryption may not be necessary, and indirectly satisfy the confidentiality.

4.2 Integrity

Integrity is designed to protect the merchant. Because the merchant only participates in the beginning step 1 and the step 6 of the trade procedure shown in figure 2. The merchant may be afraid that the price of the items or the items identity will be changed. Let the merchant add MAC which is made by hash function on the transmission message. In after procedure among transmission, the MAC needs to be transmitted without changed until the message transmitted back to the place of merchant.

4.3 Authentication

The Authentication has difference opinion because of the participant. From the view of the consumers, consumers add digital signature on the transmitted message to the bank, in order to confirm that no other people can start the trade request without consumer's permission. According to the digital signature added by the bank consumers can also authentication the bank. From the view of the bank, confirm the trade is authorized by consumers using the digital signing of consumers. From the view of the merchant, because the received receipt has digital signature of bank, the merchant can confirm the bank had already paid the

amount.

4.4 Non-Repudiation

The non-repudiation also finishes through the mechanism of digital signature. By digital signature, consumers can't deny the trade, the operator can't deny authentication consumers and pass trade request made by consumers. The bank can't deny finishing consumers' payment. The trade is invalid unless there is assurance of digital signature.

4.5 Privacy

The protection of privacy is the key of this article. About trade account which includes the credit card number and bank account, the merchant has no way to receive trade account because trade message is transmitted by consumers. In addition, the consumer is authenticated by using digital signature, and trade account of consumers is never transmitted in the network. About shopping habit, the transmitted message in the network only has the trade identity and MAC, the bank and operator only know the identity without meaning. Therefore, the system can protect privacy of consumers.

5. CONCLUSION

The proposed system in this article puts emphasis on three parts, familiarity, privacy, and non-repudiation. First, the trade behavior does not change very much. Instead of using credit card, consumers use the mobile device with them to do trade activity. Hence, consumer will just perceive the change of trade device. The mobile device, e.g. handset, is familiar to consumers. Secondly, consumer's privacy can also be protected. The merchant, operator, and bank can't accumulate shopping habit of consumers. Third, the proposed system using digital signature to protected the trade non-repudiation. Table 4 shows the comparison of relevant payment system research in the past.

Table 4. Systematic comparison of payment

	Encryption system	Digital signature system	Account protected system	The proposed system
Confidentiality	⊙	⊙	⊙	⊙
Non-repudiation	×	⊙	×	⊙
Account protect	×	×	⊙	⊙
Trade privacy	×	×	×	⊙

(⊙: Offer, ×: Does not offer)

The proposed mobile payment system satisfies not only the payment security and non-repudiation requirement in the traditional electronic trade but also the protection of consumer's privacy which is paid attention gradually now. In this system, the merchant, operator, and bank

can't accumulate shopping habit of consumers.

Perhaps someone will query that this model needs consumers to transmit trade message to the operator and this step will increase the transmission burden of the network. In fact, under the primitive transmission structure, consumers must transmit the trade message to the merchant. In our design, the message is just transmitted to operator instead. Hence, the change will not increase the transmission burden of the network.

REFERENCES

- [1] Welch, D., Lathrop, S., "Wireless Security Threat Taxonomy", *2003 IEEE Workshop on Information Assurance United States Military Academy*, pp76–83, West Point, NY, Jun. 18-20, 2003.
- [2] Soliman, H. S., Omari, M., "An Efficient Application of A Dynamic Crypto System in Mobile Wireless Security", *Wireless Communications and Networking Conference 2004/IEEE Communications Society*, Vol. 2, Mar. 21-25, 2004.
- [3] Kungpisdan, S., Srinivasan, B., Le, P. D., "A Secure Account-Based Mobile Payment Protocol", *International Conference on Information Technology: Coding and Computing (ITCC 2004)*, Vol. 1, pp35–39, Apr. 5-7, 2004.
- [4] "SET Specifications", <http://www.setco.org>, 1997.
- [5] Fourati, A., Ben Ayed, H.K., Kamoun, F., Benzekri, A., "A SET Based Approach to Secure the Payment in Mobile Commerce", *Proceedings of the 27th Annual IEEE Conference on Local Computer Networks (LCN.02)*, pp136–137, Nov. 6-8, 2002.
- [6] Howell, R., "WAP Security", *Concise Group Ltd.*, <http://www.vbxml.com/wap/articles/wapsecurity/default.asp>
- [7] WAP Forum, "Wireless Transport Layer Security Specifications", Version Apr. 2001, <http://www.wapforum.org>, April, 2001.
- [8] Herzberg, A., "Payments and Banking with Mobile Personal Devices", *Communications of the ACM*, Vol.46, No.5, 2003.
- [9] Sue, W.H., "Adaptive payment system for mobile environment", *Dissertation submitted to Department of Electrical Engineering National Taiwan University*, 2003.
- [10] [Rubin, A. D., Wright, R. N., "Off-line generation of limited-use credit card numbers", *AT&T Labs – Research*, 2002.
- [11] Ateniese, G., "Efficient verifiable encryption (and fair exchange) of digital signatures", *6th ACM conference on Computer and communications security*, 1999.
- [12] Rivest, R., "The MD5 Message Digest Algorithm", *Request for Comments (RFC) 1321*, Internet Activities Board, Apr., 1992.
- [13] National Institute of Standards and Technology, "Secure Hash Standard", *Federal Information Processing Standard Publication 180-1*, U.S. Department of Commerce, Apr., 1995.

Physicians' Opinions of an SMS Professional Medical News Service: Insights from an SMS Survey

Shengnan Han¹, Pekka Mustonen², Matti Seppänen², Markku Kallio²

¹Turku Centre for Computer Science (TUCS) & Åbo Akademi University,
Åbo/Turku, 20520, Finland

²The Finnish Medical Society Duodecim, Helsinki, 00101, Finland
shengnan.han@abo.fi; {pekka.mustonen, matti.seppanen, markku.kallio}@duodecim.fi

ABSTRACT

This paper is designed to investigate physicians' opinions regarding an SMS professional medical news service in the Finnish health-care sector. A survey using SMS mobile technology was conducted on March 5, 2003. 259 out of 685 responded within 2 days, and 90% came within 6 hours after the survey was sent out. The response rate was 38%. Findings from this simple SMS survey showed that physicians had positive perceptions of the SMS news service. Nearly 60% of the respondents have used the service. Some of the answers included spontaneous feedback about the service, which revealed valuable comments and suggestions regarding the further improvement of the service. The SMS survey as a new data collection method needs academic attention. Future research is briefly discussed.

Keywords: SMS medical news service, physicians' acceptance, SMS survey, time-saving, cost-effective

1. INTRODUCTION

Mobile commerce (m-commerce) has been an important research focus in recent years. Generally, m-commerce is defined as the extension of electronic commerce (e-commerce) from wired to wireless computers and telecommunications, and from fixed locations to anytime, anywhere and anyone [17], i.e., the use of mobile technologies and devices to provide, sell and buy convenient, personalized, and location-based services.

Text-based technology or Short Messaging Service (SMS) is one of the underlying technology platforms for m-commerce. In a comparison with wireless web-based technologies, i.e., another m-commerce platform, SMS has a simple user interface and is supported by most mobile phones. Recent years have seen the adoption of SMS worldwide in many commerce sectors, e.g., news, weather forecasting, retail, entertainment etc. It is also considered to be the most convenient and cheap communication technology used by normal consumers [23].

The rapid diffusion of SMS has also inspired some applications in the health care industry, for example, SMS professional medical news service. Medical knowledge is changing constantly. It is not easy for physicians to keep their knowledge and information up to date to help in their patient care and patient management efficiently on the one hand, and to maintain the level of their professional competence on the other [16]. An SMS professional medical news service might help physicians keep their knowledge up to date and to provide some hints of recent medical development trends and new discoveries. The objective of this paper is to investigate physicians' acceptance of

an available SMS professional medical news service that is currently implemented in the Finnish health care sector. First, we present the theoretical background underlying the study. We describe the mobile service in detail in section 3. In section 4, we present the study context and the method we used for data collection. In section 5, the results are reported. Discussions and conclusions are at the end.

2. PHYSICIANS ADOPTION OF TECHNOLOGY

Users' perceptions of and intentions to adopt information system (IS) and the rate of diffusion and penetration of technology within and across organizations are two important foci in IS research [e.g., 22]. They are understood to represent or stand for the essential aspect, property or value of the information technology [21]. It is generally accepted that the usage of information systems at work could increase employees' productivity in their working time, and improve individual and organization performance. System usage is an important dimension to measure IS success [9] [10]. In particular, physicians adoption of IS in the health care is aimed to improve the health quality of the human beings. In the past several decades, the conclusions of many studies based on different theoretical approaches and research methods, have approved and confirmed that usefulness of a supportive medical IS is very important in influencing a physician's decision making of using it. For examples, Chau and Hu [6] [7] based on classical IS adoption models [e.g., 8, 20] and collected data from a self-administered questionnaire to investigate physicians' adoption behavior of technology. They found that usefulness is a single factor determines a physician's adoption of telemedicine technology. They argued that

physicians are professionals and might exhibit considerable differences in general competence, adaptability to new technologies, intellectual and cognitive capacity and the nature of their work. Therefore, the ease of use of the technology might weaken as the competency of the user increases. Jayasuriya [15] conducted his research in a similar style. He indicated, health professionals were willing to use technology in their jobs when they perceived it to be useful for their performance. But the study results from Han et al., [14] presented that in the early stage of using mobile technology, ease of use still has strong effects on a physician's behavior intention towards the technology. Many studies were conducted by utilizing qualitative methodologies revealed similar results. For instances, Mayer and Piterman [18] studied the attitudes of Australian general practitioners to evidence-based medicine by using a focus group method. They concluded that physicians' acceptability was influenced by relevance (usefulness) to general practice, as well as the local contextual and patients factors. By using a prototype simulation study 1 week in a hospital in Germany, Ammenwerth et al. [1] successfully found out that in order to meet the diverse requirements of different professional groups, the developer needs to design "multi-device mobile computer architecture". Therefore, the usefulness of the mobile information and communication system in clinical routine is more likely to be improved.

Obviously, researchers have adapted different methods to study physicians' adoption of technology. The selection of different research methods is incumbent of individual IS researcher depending on the problem he or she aims to solve. As suggested by Moody and Buist [19], the real question is not whether the research method is appropriate *per se*, but whether it is appropriate to answer the question being asked.

3. MOBILE SERVICES AND THE SMS PROFESSIONAL MEDIAL NEWS SERVICE

Carlsson and Walden [3, 4, and 5] provided a conceptual framework for mobile services from three perspectives: the customer, the producer and management. Mobile services viewed from different perspectives have different distinguishing elements. These embedded elements actually describe what the mobile services will be. Here we only focus on the customer perspective. From this perspective, mobile services are durable technological products and technology-based services where repeat purchases are sparked off by an increase in user-perceived functionality. Seen from the perspective, the necessary distinguishing elements of mobile services are:

- 1) Flexibility, mobile services should be available anywhere, at anytime and anyhow;
- 2) Value adding, mobile services should improve productivity, they should be adaptive

to localization and they should be sensitive to customer personalization ;

- 3) A mobile technology basis, mobile services should use innovative and distinguishing features of mobile technology to enhance the quality of life (e.g. messaging, entertainment, education, information, privacy, etc)

These elements are key factors to ensure the quality and possible success of mobile services.

Duodecim Publishers Ltd. owned by the Finnish Medical Society Duodecim, has developed and provided the SMS professional medical news service (interchangeably referred to as the SMS news service or the service in the paper) to physicians currently involved in the Finnish health care sector, including hospitals, health care centers and private doctors etc. The news team at Duodecim Publishers Ltd maintains the service. The physicians subscribe the service to a specific number. After the service is issued, the physicians will get one piece of news per day. Currently, they do not need to pay for the service because of the support of Pfizer Finland Ltd.

Due to the limitations of SMS technology, the service does not include any pictures. A strong message from the developer mentions that the adoption of Multi-Media-Message (MMS) is a must in the near future. The pictures might be included as well.

4. STUDY CONTEXT AND DATA COLLECTION

The Finnish Medical Society Duodecim is a leading medical knowledge and information provider in Finland. It has placed much effort into improving the quality of medical knowledge. It has also adapted new technologies to distribute that knowledge to physicians, for instance, CD-ROM, Intranet, the Internet and wireless technology [11, 12, and 13]. The SMS professional medical news service started in 2002.

In order to investigate the physicians' acceptance and their feedback of the service, a survey was conducted on March 5, 2003. The survey was sent to the target group (685 physicians, among those, 259 subscribers to the service) as an SMS message, and the answers were gathered via the same media.

Concerning the limitations of SMS technology, the survey had only one question, i.e., "how do you perceive the SMS medical news service by Duodecim?" with three pre-defined answer alternatives, they were:

- a) "The service is really good and useful."
- b) "It is ok, but needs to be improved."
- c) "I haven't subscribed to the service."

The question was sent out on March 5, 2003 at 12:52. Totally 259 answers out of 685 were received within 2

days, a response rate of 38%. Two responses were excluded from the data analysis due to empty replies. In case of the 259 have subscribed to the service, the response rate from the subscribers was extremely good, over 50%. Some of the answers included spontaneous feedback about the service, which was very valuable. The majority of answers (>90%) were gathered within 6 hours of sending out the question. Before mid-night of March 5, we got 243 answers, only sixteen was sent back on March 6.

5. RESULTS

We did a basic summary report of our data. Since some answers provided feedback about the services other than our pre-defined alternatives, we consider those responses as D. Results are summarized in Figure 1.

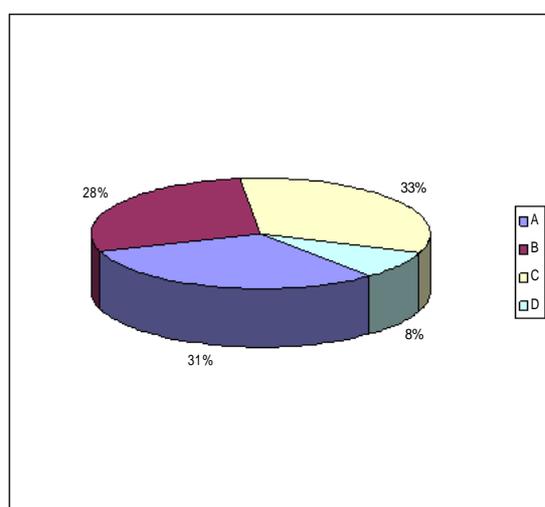


Figure 1 Survey results

Of 257 usable answers, thirty-one percent (n=79) perceived that the service was really good and useful. Seventy three, i.e., 28% thought improvement of the services was necessary. Thirty-three percent of the physicians had not yet subscribed to the service. We got 22 extra answers that were not covered by our pre-defined answers, which we counted as D. Below we present the details of the feedback spontaneously provided with the three alternatives and other comments categorized as D.

I Arguments with A

Physicians who replied with the A alternative really liked the service and thought it was very useful for their work. They also showed a positive future intention towards the service. Such as a physician wrote,

"[The service] is very useful and I would like to receive the messages in the future..."

Besides acknowledging the usefulness of the SMS news service, physicians are still expecting information from traditional channels, a physician replied,

"The messages are excellence for me and I would be happy to receive other news letters as well."

I Arguments with B

Most of those who selected B did not specify very much about how to develop the system. Several physicians did give suggestions. A physician mentioned that

"Occasionally one gets news from a short message. [It is] a different picture of the matter compared with the original text [e.g., newspapers]."

This message indicated that, as an emerging information distribution channel, mobile SMS was expected to be compatible with other traditional media, such as the texts published in newspapers or on the Internet. On the other hand, physicians also welcomed the SMS news, *"I have during weekdays several times had the time to read the same news on the Internet, but it's good to have repetitions"*.

Three physicians suggested that the SMS news might include more extensive sources or specialized fields.

"... [the news might come] along with Internet references or more extensive sources"

"...Develop the messages for specialized fields, i.e. information to the correct address..."

"... More accurate sources of references could make it easier to find the article (e.g., about the disadvantage of epilepsy medication)"

Those suggestions shed light on how to improve the content of the services. The reference sources are quite demanding that help physicians go further to check the latest medical information. Physicians also need information from their specialized fields that are better suited to meet their personal requirements on specific knowledge they want to acquire.

Some physicians have pointed out that the SMS news should be longer, the shortness of SMS was considered a drawback. Comments came like,

"The messages could be even longer".

"Apparently the shortness of SMS causes limitations"

I Arguments with C

The feedback came with alternative C provided our several reasons why physicians did not subscribe to the services. The first and obvious reason was they didn't need the SMS news service, e.g,

"Not in use and not needed"

"Not in use and I don't want it."

The second reason was some physicians still stick to other distribution channels, such as newspapers or the Internet

"I haven't followed the text message news. I read newspapers."

"It is easier to read from the internet"

The third reason was that technical problems hindered physicians' usage of the SMS news service. Complaints came like,

"I have not received any news even though I have subscribed";

"I have tried to get the service without success..."

"Not in use since I don't know how..."

The fourth reason was due to the small screen of the mobile device, *"I would prefer to read them as emails due to the small screen of my mobile phone...probably better with the communicator."* The size of the screen is not comparable with PC-based media apparently. Some physicians also lacked the time to subscribe to the service, *"I'm in principle interested, but haven't had time to pick up the instructions"*. Physicians are usually very busy. Lack of time is frequently cited as a reason for them not adopting new information systems [2].

Besides those negative comments, we also found positive intentions from some non-adopters. Such as,

"I will activate"

"I would like to have in use..."

"Reading news as SMS can in principle be feasible".

1 Arguments collected from D

A deep analysis of the feedback we classed as D was conducted as well. We found out that there were three main strong suggestions from the physicians.

1) The service still had some bugs, problems with receiving and opening the service were common complaints. For examples,

"None of the text message news that I ordered have been activated, so I have not received any news for the time beings";

"I have not received any news even though I got the confirmation of order";

"For some reason, I have not received news for a long time";

"I did according to instructions, didn't receive messages";

"How is it activated?"

These messages indicated that the developer and the service content provider have to pay much attention to the technical problems associated with the service. The growing complaints of the difficulties in activating and receiving the service might become a severe barrier for physicians to continue their adoption of it.

2) The current length of the message (160-200 characters) was considered both a limitation and an advantage. We got free -form feedback like,

"160 characters is quite limited amount. But on the other hand, one doesn't bother to read too long stories""During office hours short and concise..."

Obviously the limited length has made the news concise. Physicians might save their time for reading the news in a busy working day.

3) The news service is quite general; more categories of specialties and more in-depth news are required. For example,

"Seems like a good idea, personally, I would like to choose those fields that interest me"

"I would wish for more information about urology..."

These messages revealed that physicians demand more services that meet their "personal" needs. The current service is targeted to all physicians as one uniform group. The developer has to take "personalization" into consideration for the further improvement of the service.

6. DISCUSSION AND CONCLUSION

This paper is intended to investigate physicians' acceptance and opinions regarding the SMS professional medical news service in the Finnish health-care sector. The findings from this simple SMS survey showed that physicians had positive perceptions of the service. Those responses located in the alternatives of A and B indicated that, generally, the ones (31%) who were happy with the service really liked it, and an almost equal amount, i.e., 28% thought the service should be improved. The total answers that went to A and B reached 59%, nearly 60%. In general, the SMS professional medical news service is useful. User adoption theories assert that usefulness is the main driver of individual adoption of a specific information technology. The perceptions of usefulness of the SMS medical news by physicians demonstrate strong possibilities of continuous usage in the future.

Arguments came with pre-defined alternatives and those categorized as D overlapped in some cases. In brief, they highlighted several important issues concerning the improvement of the service in the future. Firstly, technical problems have to be solved immediately. It constitutes a barrier to physicians' usage of the service continuously, or hinders their subscriptions. Possible guidance on how to open and use the service should be improved as well. Ease of use is always considered a crucial driver of the success of information system implementation. Mobile technology is not an exception [14]. Secondly, the length limitation of SMS technology has made not only the news concise, which considered a time-saving way to read it, but also made it impossible to include more information in one message. With the development and diffusion of MMS,

technically the news items could easily be made longer. Thirdly, “personalization” is demanded by the physicians. The general topics adapted in the current service did not take use of the advantages of mobile technology, i.e. personalized services to mobile users. Therefore, it might undermine the value added by the SMS medical news service. Personalized service for the physicians’ news needs will enhance their knowledge what is highly relevant to their specialties, and provide the information that they desire most. Fourthly, as an emerging technology, SMS is competing with traditional channels of distributing medical news. The traditional channels, e.g., newspapers and the Internet, have dominated information delivery for a long time. Currently, SMS is a complementary channel, but far from the dominating one. A good thing is that it is warmly welcomed by most of the physicians. It has strong growth potentials in the future. It comparatively adds more value to physicians’ work in terms of flexibility and personalization or localization.

Nearly one third of the respondents had not yet subscribed to the SMS professional news service. Attention has to be paid to the reasons why they did not subscribe to the services, in order to increase the diffusion speed. Compared with the aforementioned important issues for the improvement of the service, we find that strong effort of improvement might remove the negative “reasons” that impede some physicians’ subscription to the service.

The results of the paper were from an SMS survey. It is a new data collection method which is time-saving and cost effective. We got 38% responses within 2 days, and 90% came within 6 hours after the survey was sent out. This new survey method needs our attention. It is a method welcomed by practitioners. The simple survey provided the answers in a very short time. The service developer and content providers received “instant” results that help them investigate the problems and find solutions almost at the same time. The SMS survey does not like traditional surveys that usually take months or years; it is a new attempt to do the survey in hours or days with aid from new mobile technology. It also reduces the cost of the survey. Mail surveys are usually very expensive, as are Internet surveys that have a monthly fee. The main weakness of the SMS survey is obviously its length limitation; we can not have many questions in the survey. But it helps us to concentrate on the most important message that the practitioners want to know immediately. Such instant feedback might represent high business value for practitioners.

In the future, as MMS is on its way to replacing the current SMS, technically, it is easy to overcome this limitation, and we might include more and longer questions in a MMS survey. With the support from

Pfizer Finland Ltd., such a survey is planned to be conducted at the autumn in 2004.

ACKNOWLEDGMENT

Our warmest thanks go to Pfizer Finland Ltd. for its support for conducting the survey. The first author also would like to reserve thanks to Ms Stina Störling-Sarkkila for her support in translating the Finnish data into English.

REFERENCES

- [1] Ammenwerth E. et al (2000): Mobile information and communication tools in the hospital, *International Journal of Medical Informatics* 57 (2000) 21-40
- [2] Berg M. (1999): Patient care information systems and health care work: a sociotechnical approach, *International Journal of Medical Informatics* 55 (1999) 87-101
- [3] Carlsson, C. and Walden, P. (2001), The Mobile Commerce Quest For Value-Added Products And Services. In proceedings of *SSGRR 2001*, L’Aquila, Italy, August 6-12, 2001.
- [4] Carlsson, C. and Walden, P. (2002a), Mobile Commerce: A summary of quests for Value-added products and services. In the proceedings of *15th Bled Electronic Commerce conference, e-Reality: Constructing the e-economy*, Bled, Slovenia, June 17-19, 2002
- [5] Carlsson, C. and Walden, P. (2002b), Extended quests for value-added products and services in mobile commerce. In the proceedings of *International conference on decision making and decision support in the Internet Age*, July 4-7, Cork, Ireland
- [6] Chau, P.Y.K. and Hu, P.J. (2002 a), ‘Investigating healthcare professionals’ decisions to accept telemedicine technology: An empirical test of competing theories’, *Information & Management*, Vol. 39, No. 4, pp. 297-311
- [7] Chau, P.Y.K. & Hu, P.J. (2002 b), ‘Examining a model of information technology acceptance by individual professionals: An exploratory study’, *Journal of Management Information Systems*, Vol. 18, No. 4, pp. 191-229
- [8] Davis, F.D.; Bagozzi, R.P. and Warshaw, P.R. “User Acceptance of Computer Technology: A Comparison of Two Theoretical Models,” *Management Science* (35:8), August 1989, pp. 982-1003.
- [9] DeLone, W., and McLean, E. (1992): Information systems success: The quest for the dependent variable. *Information Systems Research*, 3(1), 60-95
- [10] DeLone, W., and McLean, E. (2003), The DeLone and McLean Model of Information Systems Success: A ten-year update. *Journal of Management Information Systems/Spring* 2003, Vol. 19 No.4, p9-30

- [11] Han S., V. Harkke, P. Mustonen, M. Seppanen and M. Kallio (2004 a): Physicians' perceptions and intentions regarding a mobile medical information system: some basic findings, Proceedings of 15th IRMA international conference, New Orleans, USA, 23-26 May 2004
- [12] Han S., V. Harkke, P. Mustonen, M. Seppanen and M. Kallio (2004b): Professional mobile tool: A survey of Physicians' perceptions of and attitude towards a mobile information system, Proceedings of 5th GITM conference, San Diego, USA, 13-15 June, 2004
- [13] Han S., V. Harkke, P. Mustonen, M. Seppanen and M. Kallio (2004c): Mobilizing medical information and knowledge: some insights from a survey, Proceedings of ECIS 2004, Turku, Finland, 13-15 June, 2004
- [14] Han S., P. Mustonen, M. Seppanen and M. Kallio (2004d): Physicians' behavior intentions regarding a mobile medical information system: an exploratory study. Proceedings of AMCIS2004, New York City, New York, 5-8 August, 2004
- [15] Jayasuriya R. (1998): Determinants of microcomputer technology use: implications for education and training of health staff, *International Journal of Medical Informatics* 50 (1998) 187-194
- [16] Jousimaa J. (2001): *The clinical use of computerised primary care guidelines*, Doctoral Dissertation, University of Kuopio, Finland.
- [17] Keen, P and Mackintosh, R. (2001), *The freedom economy: Gaining the M-commerce edge in the era of the Wireless Internet*, Osborne/Mcgraw-Hill, Berkeley; CA, USA.
- [18] Mayer J. and Piterman L. (1999): The attitudes of Australian GPs to evidence-based medicine: a focus group study, *Family practice* 1999, 16: 627-632
- [19] Moody, D.L., and Buist, A. (1999) Improving Links between Information Systems Research and Practice-Lessons from the Medical Profession, in proceedings of 10th ACIS, 645-659.
- [20] Morre, G.C., and Benbasat, I. (1991), Development of an instrument to measure the perception of adopting and information technology innovation. *Information systems Research*, 2(3), 192-223
- [21] Orlikowski, W.J and Iacono, C S (2001), Research commentary: Desperately seeking the "IT" in IT research – A call to theorizing the IT artefact, *Information Systems research*, Vol.12, No.2, June 2001, p121-134.
- [22] Straub, D.W., M. Limayem & E. Karahanna-Evaristo (1995), 'Measuring System Usage: Implications for IS Theory Testing', *Management Science*, Vol. 41, No. 8, pp. 1328-1342
- [23] Xu, H., Teo, H.H and Wang., H. (2002): Foundations of SMS Commerce Success: Lessons from SMS messaging and Co-opetition, In Proceedings of 35th HICSS, 2002

Security Measures in Mobile Commerce: Problems and Solutions

Sanwar Ali¹, Waleed Farag², Mohammad A. Rob³

¹Department of Computer Science, Indiana University of Pennsylvania, Indiana, PA 15705, USA
sanwar@iup.edu, Fax: 724-357-2724

²Department of Computer and Systems Engineering, Zagazig University, Zagazig, Egypt
farag@cs.odu.edu, Fax: +20-55-2304987

³Management Information Systems, University of Houston- Clear Lake, Houston, TX 77058, USA
rob@cl.uh.edu, Fax: 281-283-3951

ABSTRACT

Due to the advent of the Internet, electronic business transactions have exploded around the globe. Along with the Internet, wireless technology has exponentially developed as well. Today, new technologies that allow mobile (cellular) phones and other handheld devices to access the Internet have made wireless business transactions possible. This phenomenon is known as mobile commerce or M-Commerce. It has been predicted that the number of mobile phones connected to the mobile Internet will exceed the number of Internet-connected PCs before 2007. The mobile phone will therefore become the most prevalent device for accessing the Internet. Several industry analysts predict that M-commerce will constitute a multibillion dollar business by 2005. However, M-Commerce brings new challenges in providing information security as information travels through multiple networks often across wireless links. What must be done to secure financial transactions via mobile commerce? Generally speaking, M-Commerce creates more security concerns than traditional E-Commerce. In this paper, security measures in M-Commerce, wireless security, and the application of cryptography for key generation, authentication, digital signature and digital certificate are discussed.

Key words: M-Commerce, wireless security, elliptic curve cryptography, WLAN security, Bluetooth

1. INTRODUCTION

Mobile commerce, commonly known as M-Commerce, 'is concerned with the use, application and integration of wireless telecommunication technologies and wireless devices', such as Internet-enabled mobile phones, personal digital assistants (PDA), palmtops, laptops, and pagers, commonly known as handheld wireless devices (HWD). The discipline of M-Commerce includes reference to the infrastructures and electronic technologies necessary for wireless data and information transfer in the form of text, graphics, voice, and video. M-Commerce is a subset of electronic commerce where the Internet-enabled HWDs and wireless networking environment are necessary to provide 'location independent connectivity'. It is predicted that M-Commerce services would be the next biggest growth area in the telecommunications market, representing the fusion of two of the current consumer technologies: wireless communications and E-Commerce [7].

Mobile communications is now considered a relatively mature technology with the transition from second to third generation (3G) wireless technology and with its high consumer acceptance. 3G wireless technology provides a variety of services and capabilities in addition to voice communication, such as multimedia data transfer, video streaming, video telephony, and full Internet access. 3G mobile phones provide high-speed data transfer and are designed to support large numbers of users more efficiently allowing for future expansion in user capacity. This technology was first introduced in Japan in 2001 and then spread in Europe and USA in

2002. In Japan mobile commerce has been a huge success, accounting for over \$400 million in revenues each year [8]. In 2003, there were about 800 million Web-enabled cellular phones in use worldwide. By 2004, the worldwide number of Web-enabled mobile phones is projected to rise to one billion. The M-Commerce market may rise to \$200 billion by 2004 (Gartner Group). It has been predicted that the number of mobile phones connected to the mobile Internet will exceed the number of Internet-connected PCs before 2007. The mobile phone will therefore become the most prevalent device for accessing the Internet and M-Commerce. Several industry analysts predict that mobile commerce will constitute a multibillion dollar business by 2005. However, once again, the question of security arises. What must be done to secure financial transactions via mobile commerce? Secure mobile commerce will be a key area within the expanding functionality of mobile phones. Generally speaking, M-Commerce creates more security concerns than traditional E-Commerce.

2. APPLICATIONS OF M-COMMERCE

There is a wide variety of applications of M-Commerce. For example, Smart cards with an embedded integrated microchip can be used as prepaid phone cards, ATM cards, or public transportation cards. They can be biometrically enhanced to include voice recognition, iris and face scans, and finger print authentication. In France, some 35 million Smart cards are in circulation and every year they process over three billion transactions [4]. Wireless banking refers to purchasing over Internet-enabled HWDs like wireless application protocol (WAP)

phones or PDAs. Interactive TV is enhanced TV where additional content is added to an existing broadcast format that the viewer can query, request or even interact live with the program. It has reserved channels and bandwidth for data applications such as weather, news, games, or commerce. It also offers services like Video on Demand (VOD) or Personal Video Recording (PVR). Companies are providing facilities to track stocks on HWDs. Aspiro, a Swedish company, allow its customers to check stock prices or look at their portfolios and even trade using WAP phones or PDAs. Accessing information using WAP mobile phones and PDAs is significantly becoming popular for business-to-business (B2B) and business-to-consumer (B2C) applications. Mobile and wireless technologies have been an integral part of defense and military ever since these technologies were available. In light of growing concern with cyberterrorism, security has undoubtedly become the single most important issue because many U.S. intelligence services use mobile technologies and wireless networks for communication and commerce.

3. M-COMMERCE SECURITY CONCERNS

M-Commerce is bringing together two technologies, wireless communication and traditional E-Commerce, with a history of security problems. Coupled with the convergence of voice and data communications, interconnection with external data networks and issues surrounding the transactions themselves, the potential risks are very high [7]. There are three basic security components in M-Commerce: (a) *Transaction*: protecting the transaction parties and their data by providing an acceptable level of security, (b) *Information*: protecting valuable and sensitive information about customers, and (c) *Infrastructure*: protecting the network infrastructure from attack.

Mobile 3D is Visa International's new global secure specification that is expected to ensure the security of Internet payments made over mobile devices [14]. It was developed in conjunction with some 15 industries, including Ericsson, Motorola, and Oracle Mobile, and was launched in September 2001. It extends payment authentication initiatives into M-Commerce, enabling the Visa card issuers to validate the identity of their cardholders in real time, ensures that payment data sent over open networks is not compromised and allows consumers to actively protect their Visa accounts from unauthorized use, and supports global interoperability, enabling consumers to have a consistent and seamless experience regardless of the method or device being used to access the Internet. A number of Visa M-Commerce programs are currently underway worldwide to test the validity of M-Commerce payment solutions and raise consumers awareness. In Asia, Visa has partnered with Hutchison Telecommunications and Dao Heng Bank to develop a mobile payment service using Mobile 3D Secure. In Europe, Visa has made a strategic alliance with Omnitel Vodafone, while in the US, Visa and Sprint

are working together to help facilitate secure mobile payments.

A satisfactory level of security is required for the successful deployment of HWDs. To a certain extent, it seems reasonable to utilize solutions used for the wired environment in the case of the wireless environment. However, this approach is not always feasible because of the differences between the wired and wireless environments. For example, because of the hardware limitations of the HWDs, no large routing tables can be maintained on these devices, thus increasing the risk of a denial-of-service attack. Furthermore, wireless communications make physical eavesdropping almost undetectable [10]. Similar to wired communication, wireless communication also needs three basic security requirements: (i) confidentiality- information is disclosed only to legitimate entities or processes, (ii) integrity- unauthorized modification of information is prevented, and (iii) availability- authorized entities can access a service provided they have appropriate privileges. The mechanisms suitable to the HWD are discussed here.

M-Commerce needs several layers of security: (i) device security, (ii) language security, (iii) wireless security, and (iv) cryptographic security.

3.1 Device Security

Within the design of mobile devices, there are a number of high quality security features. The most important of these are: (i) a built-in password mechanism which will lock after several mistyped attempts and (ii) an industry approved, tamper-proof smart card, known as Subscriber Identification Module (SIM) card.

The SIM card and the mobile device are always stored together and the device is an every day utility object that is easily lost or stolen. Time-out and key-locks are often not used on phones. This means that as long as the phone remains turned on the strong password system will be bypassed. All WAP data, in some popular handsets, is stored in the phone's memory, not in the SIM; this will include login and password information. These features certainly diminish the security of the mobile phone. The SIM card used in mobile devices are *de facto* micro-processor and they can be used to facilitate mobile commerce. Gemplus SIM cards features a digital signature and public key encryption [11] and the technology is embedded in the card. In May 1999, Motorola, jointly with Identix, a biometrics company, developed fingerprint scanning devices, called the DFR 300 that is 4.5 millimeters thick [2]. This scanning device can now be incorporated into the HWDs.

3.2 Language Security

If special purpose M-Commerce software, such as a stock trading application, is to be deployed on mobile

devices, then Java is the recommended language to be used as the deployment language on the HWDs. By using Java, the amount of software that needs to be changed in order to adopt the application to various mobile platforms is minimized. Feasible Java execution environments are available for PDAs, Smart phones, Communicators (such as Symbian), laptops, and other platforms. Maffeis [6] also recommended using server side Java technology, such as the Jave-2 Enterprise Edition (J2EE) platform, in the data center. This allows for shorter time-to-market and avoids vendor lock-in.

3.3 Wireless Security

3.3.1 WAP Security

WAP (Wireless Applications Protocol) is 'an open, global specification that empowers mobile users with wireless devices to easily access and interact with information and services instantly.' WAP is currently the only publicly available solution for wireless communication and enables M-Commerce where Internet data moves to and from wireless devices. WAP-enabled phones can access interactive services such as information, location-based services, corporate information and interactive entertain-ment. WAP is targeted at various types of HWD and Bluetooth enabled mobile phones.

WAP 1.x security uses the Wireless Transport Layer Security (WTLS) protocol. This protocol is the WAP equivalent of Secure Socket Layer (SSL) and it provides authentication, encryption and integrity services. WTLS has three levels, all have privacy and integrity: (i) *Class-I* has no authentication (anonymous), (ii) *Class-II* has server authentication only, and (iii) *Class-III* has both client and server authentication. WTLS supports some familiar algo-rithms like Diffe-Hellman, RC5, SHA-1, and IDEA. It also supports some trusted methods like DES and 3DES, but it does not support Blowfish and PGP [9].

Since Web- and WAP-based protocols are not directly interoperable, a component known as the WAP gateway is needed in order to translate Web-based protocols to and from WAP-based protocols. The WAP gateway is a software that runs on the computer of the Mobile Service Provider (MSP). Thus sensitive information is translated into original unencrypted form at the WAP gateway [5]. This problem is known as WAP gap.

Public key cryptography (PKC) is used to exchange a symmetric or private key using certificate and then all transmission is encrypted. A short key size of 40 bits is used because of power limitation. A tamper-proof compo-nent, know as WIM (Wireless Identity Module) is designed as part of the WAP architecture to store private data, such as key pairs, certificates, and PIN numbers within the mobile device. In practice, a WIM is imple-mented using a smart card. Wireless Markup

Language (WML) is used in WAP 1.x technology. Figure 1 shows WAP gap model.

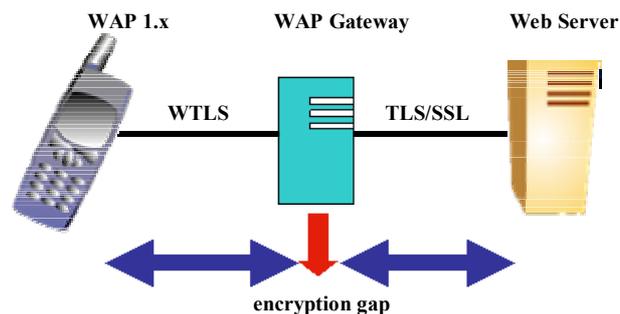


Figure 1 WAP Gap Model (not a full end-to-end security)

WAP 2.0 security uses TLS (Transportation Layer Security) instead of WTLS due to requiring end-to-end security with all IP based technology in order to overcome the WAP gateway security breaches. It is a Public Key Infrastructure (PKI) enabling protocol that provides the services such as authentication by using digital signatures and public key certificates, confidentiality by encrypting data, etc. This protocol uses RSA, RC4, 3DES, and SHA-1 algorithms for encryption. Wireless PKI (WPKI) is released for the first time. Figure 2 shows the WAP proxy model.

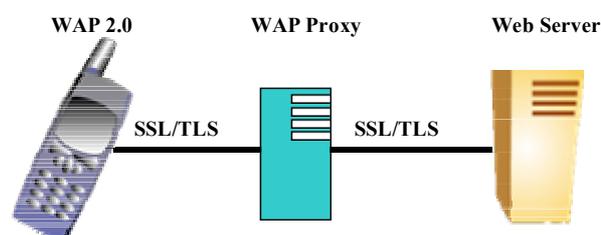


Figure 2 WAP Proxy Model (end-to-end security)

3.3.2 PKI/WPKI

PKI systems and WTLS are at the heart of today's mobile security technology. In a WAP environment WTLS must be translated at the WAP gateway into SSL, the Internet standard. A PKI is a set of policies, processors, software, hardware, and technologies that use PKC and certificate management to secure communication [18]. PKI's trusted services enable the secure transfer of information and supports a wide variety of M-Commerce applications. PKI must ensure the following: (i) confidentiality, achieved by cryptography, (ii) authentication, achieved by digital certi- ficates, (iii) integrity, achieved by digital signatures, and (iv) nonrepudiation, achieved by digital signatures and certificates.

PKI consists of the following components: (i) Certificate Authority (CA)- responsible for issuing and revoking certificates, (ii) Registration Authority (RA)- binding

between public key and the identities of their holders, (iii) Certificate Holders- people, machine or soft-ware agents that have been issued with certificates and can use them to sign digital documents, (iv) Verification Authority (VA, Clients)- validate digital signatures and their certificates from a known public key of a trusted CA, and (v) Repositories- stores that make available certificates.

WPKI is an optimized extension of traditional PKI for the wireless environment. 'WPKI encompasses the necessary cryptographic technology and a set of security management standards that are widely recognized and accepted for meeting the security needs of M-Commerce' [18]. WPKI applications have to work in an environment with less powerful CPUs, less memory, restricted power consumption, and smaller displays. WPKI solutions are likely to employ "network agents" that take care of some of these tasks. The HWD must at least be able to perform a digital signature function to permit the establishment of a WPKI. Network agents can perform all other WPKI-related tasks such as validation, archiving or certificate delivery. Figure 3 shows a schematic diagram of WPKI.

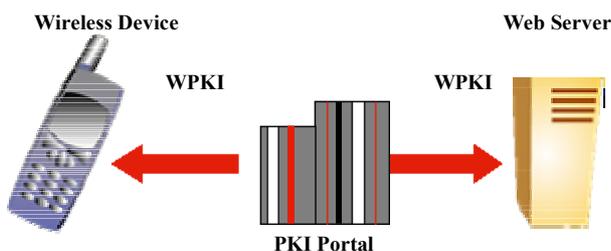


Figure 3 Wireless PKI

The private keys are stored in WIM or SWIM. Two main components of WPKI are the PKC and the key certification management. In order to perform globally, worldwide PKI legislation is required.

A number of companies, including Entrust, Certicom, RSA Security, VeriSign, and Baltimore, have announced solutions enabling the use of PKI software in a wireless environment. The biggest challenge in implementing a PKI solution for wireless deployment lies in the devices themselves. With limited bandwidth and low power, a small screen and no keyboard, the average mobile device presents a number of unique problems. Certicom developed Elliptical Curve Cryptography (ECC) that reduced the key size from RSA's 1024 bits to as few as 56 bits and made handling certificates a lot easier for low-bandwidth and low-power devices [2]. This technique can process a digital signature in a second compared to the RSA technique, which takes almost 15 seconds to process a digital signature on a Palm device. Both Nokia and Ericsson have been supporting server certificates since the middle of 2000.

3.4 Wireless LAN (WLAN) Security

3.4.1 IEEE 802.11b

The WLAN standard IEEE 802.11b provides a mechanism for authentication and encryption. It provides a maximum of 11 Mbps wireless Ethernet connections using the band at 2.4 GHz. 802.11b security features consists of security framework called Wired Equivalent Privacy (WEP). WEP is based on RC4, a symmetric stream cipher. It has a pseudo-random number generator, whose output is XORed to the data. WEP can use 40 or 128 bits key size. However, using a 128 bits key size, 802.11b throughput drops much due to heavy calculations. In August 2001, RC4 was announced to be broken and can be cracked in less than half an hour. Consequently, WEP can be broken. WEP with 40 bits key size can be broken in real time.

3.4.2 Bluetooth

Bluetooth technology, developed by Ericsson in 1998, is used to connect different HWDs and provides a method for authenticating devices. Device authentication is provided using a shared secret between the two devices. The common shared secret is called a *link key*, generated from PIN. This link key is established in a special communication session called *pairing*. All paired devices share a common link key. There are two types of link keys: (i) *unit keys* and (ii) *combination keys* [17]. The link key is a 128-bit random number.

A device using a unit key uses the same secret for *all* of its connections. Unit keys are appropriate for devices with limited memory or a limited user interface. During the pairing procedure the unit key is encrypted and transferred to the other unit. Only one of the two paired units is allowed to use a unit key. Combination keys are link keys that are unique to a particular pair of devices and they are only used to protect the communication between these two devices. Clearly a device that uses a unit key is not as secure as a device that uses a combination key. Since a unit key is common to all devices with which the device has been paired, all such devices have knowledge of the unit key. Consequently they are able to eavesdrop on any traffic based on this key.

In every Bluetooth device, there are four entities used for maintaining the security at the link level: (i) the Bluetooth device has an IEEE defined 48-bit unique address, (ii) a private authentication key which is a 128-bit random number, (iii) a 8-128 bit long private encryption key, and (iv) a random number, which is frequently a changing 128-bit number that is made by the Bluetooth device itself [12]. The security algorithms of Bluetooth are considered strong. Bluetooth standard does not use the RC4 cipher; rather it uses the E1, a modified block cipher SAFER+. No practical direct attack has been reported.

4. CRYPTOGRAPHY

The two types of cryptography currently available are (a) symmetric or secret-key and (b) asymmetric or public key cryptography (PKC). In secret-key cryptography, two devices must share their secret key in order to communicate securely. Thus two concerns arise: how to exchange the secret key securely; and if n HWDs must communicate with each other, a total number of $O(n^2)$ secret keys must be exchanged. The management of such a number of secret keys should consider the scalability issues.

In PKC, both the above problems are solved since two parties in the communication do not require exchanging any secret keys. According to this consideration, PKC seems to be the ideal candidate to enforce confidentiality. Cryptographic techniques already exist to protect data transmission over the Internet. The *de facto* cryptographic algorithm for digital signatures and for encryption of secret keys is the RSA PKC. For electronic commerce, it is a highly secure technique. Although RSA is highly secure and widely used, there are some potential problems with its use in M-Commerce: (i) in RSA the key size is large which requires a significant amount of memory for storage, (ii) decryption time increases rapidly as the key size increases, and (iii) key generation is complex and time consuming. Memory constrained devices cannot easily generate RSA keys and therefore may need to have keys generated by another system. Moreover, the computations needed to encrypt and decrypt messages using RSA is overwhelming compared to secret key cryptography.

4.1 Elliptic Curve Cryptography (ECC)

What is really needed is a public key algorithm that achieves a high level of security using short keys. Algorithms based on mathematical objects known as elliptic curves offer interesting possibilities [1]. Elliptic curve discrete logarithm problems (ECDLP) is defined as "give a base point P and the kP lying on the curve, find the value of k ". From cryptographic point of view, a new cryptographic system needs to be defined based on elliptic curves. Any standard system that relies on the discrete logarithm problem has a direct analogy based on the ECDLP. For example, Elliptic Curve Digital Signature Algorithm (ECDSA) has already been standardized. Diffie-Hellman key exchange can be easily implemented in an elliptic curve system.

Certicom promotes the use of ECC, as well as its mobile Virtual Private Network (VPN) client, and a client, which bring sophisticated security functionality to resource-constrained devices. WPKI ECC is recognized as the most optimized and therefore the best suited for supporting security in the wireless environment. The key for digital signature in ECC is 163 bits as compared to 1024 bits in other signature schemes. However, good

elliptic curves must be carefully chosen, otherwise it might be prone to various attacks. ECC is small, efficient and requires low power. Table 1 shows key sizes for ECC and RSA.

Table 1 EEC Key Size compared to RSA

ECC Key Size (bits)	Traditional RSA Key Size (bits)	Key Size Ratio
109	512	1:5
131	768	1:6
163	1,024	1:6
283	3,072	1:11
409	7,680	1:19
571	15,360	1:27

Recently, ECC has been deployed on Smartcards without coprocessors [16]. Weimerskirch *et al* [15] implemented ECC on a Palm OS device. Their study showed that the normal transaction, such as a key exchange or signature verification, can be done in less than 2.4 seconds while signature generation can be done in less than 0.9 seconds. Table 2 shows the three major industry standard PKC systems that can be considered secure, efficient, and commercially available [13].

Table 2 Three Major Industry-Standard PKC

PKC	Mathematical Problem	Algorithm
Integer factorization	Given a number n , find its prime factors	RSA, Rabin-Williams
Discrete logarithm	Given a prime n , and numbers g and h , find x such that $h = g^x \pmod n$	ElGamal, Diffie-Hellman, DSA
EC discrete logarithm	Given an elliptic curve E and points P and Q on E , find x such that $Q = xP$	EC-Diffie-Hellman, ECDSA

4.2 Problems with Elliptic Curve System

The true difficulty of the ECDLP is not yet fully understood [1]. Recent research has shown that some elliptic curves that were believed suitable for ECC are, in fact, not appropriate. For example, if the order of the base point P is equal to the prime number p then it turns out that the ECDLP can be solved efficiently. Such curves are known as anomalous curves. For a given curve and base point, it is trivial to generate public and private keys (the private key is simply a random integer k and the public key is the point kP on the curve). However, it is an extremely difficult problem to generate a suitable curve and base point in the first place.

The main problem is how to count the number of points on the curve. Having done this, it is then necessary to select a suitable base point P , which must have a large order to ensure the difficulty of the ECDLP. But the order of P must divide the number of points on the curve.

Having found the number of points on the curve, it is quite likely that a suitable base point cannot be found. Users may use random curves or special curve generating software, such as the "Elliptic Curve Generation Bureau" created by Zaxus.

In April 2000, the French National Institute for Research in Computer Science and Control announced that the 109-bit ECC key was cracked in a four-month brute-force effort using 9,500 computers by 1300 volunteers from 40 countries [3]. The amount of work required was about 50 times that required to solve a 512-bit RSA cryptosystem. The NIST has endorsed 163-bit ECC and WAP standards to be used in HWDs.

4.3 Digital Certificates and Digital Signature

In a PKC, a message is encrypted using a public key and is decrypted using a secret key. However, there is no inherent way of knowing the person who has the corresponding secret key. This is where the idea of certificates arises. Certificates confirm that the public key given in the certificate belongs to a private key held by the legitimate person, not by an imposter. To trust a certificate means to trust the party who issued the certificate, not the person for whom the certificate is issued. To protect a certificate from being modified one uses digital signatures. The message can only be created from the ciphertext by the private key holder. This provides authorization and non-repudiation. That is the basis for digital signature.

How to protect private keys? The private key is stored in a Smart card, where all crypto operations with it are performed. The Smart card access gets restricted by the use of a PIN. The place where a user's private credentials are stored is called Personal Security Environment (PSE).

5. CONCLUSION

M-Commerce security is a very crucial issue that needs further research to introduce efficient and effective solutions. In this article, various security concerns were expounded. ECC certainly appears to provide a viable alternative to RSA. There are potential advantages, especially when used in devices with limited processing capability and memory. Typical applications include M-Commerce using handheld wireless devices. There are, however, some problems and issues that are inhibiting the widespread adoption of ECC. These include (i) the real security of such systems is still not well understood, (ii) difficulty of generating suitable curves, and (iii) relatively slow signature verification. Time will tell its future.

REFERENCES

[1] Ganley, M.J. "Elliptical Curve Cryptography",

Zaxus White Paper, pp1-9, 2000.

[2] Goldman, Jeff, "Wireless Security and M-Commerce", The Feature, March 8, 2001, <<http://www.thefeature.com/article?articleid=9862>>

[3] Harrison, A., "Motorola, Certicom Ink Elliptic Crypto Deal", Computerworld, May 22, 2000.

[4] "How the French are Succeeding with M-commerce", Wireless developer Network, <<http://www.wirelessdevnet.com/channels/wap/features/mcommerce4.html>>

[5] Juul, Niels C. and Jorgensen, N. "WAP may stumble over the Gateway", 2001, <<http://webhotel.ruc.dk/ncjuul/papers/wap.pdf>>

[6] Maffeis, S. "M-Commerce Needs Middleware!", 2000, <<http://www.softwired-inc.com/people/maffeis/articles/softwired/mcommerce.pdf>>

[7] Messham, James, "M-Commerce Security", <[http://www.tdap.co.uk/uk/archive/billing/bill\(fml_0012\).html](http://www.tdap.co.uk/uk/archive/billing/bill(fml_0012).html)>

[8] "Mobile Commerce (M-commerce)", <<http://www.fiercewireless.com/topics/mcommerce.html>>

[9] Osborne, Mark, "WAP, m-commerce and security", 2000, <<http://www.kpmg.co.uk/kpmg/uk/image/mcom5.pdf>>

[10] Pietro, Robert D. and Luigi V. Mancini, "Security and Privacy Issues of Handheld and Wearable Devices", Communication of the ACM, Vol. 46(9), pp75-79, September 2003,

[11] "PKI Moves Forward Across the Globe", Wireless developer Network, <<http://www.wirelessdevnet.com/channels/wap/features/mcommerce3.html>>

[12] Vainio, J.T. "Bluetooth Security", 2000, <<http://www.niksula.cs.hut.fi/~jjitv/bluesec.html>>

[13] Vanstone, S.A. "Next generation security for wireless: elliptic curve cryptography", pp412-415, 2003, <http://www.compseconline.com/hottopics/hottopic20_8/Next.pdf>

[14] "Visa Mobile 3D Secure Specification for M-commerce Security", <http://www.cellular.co.za/technologies/mobile-3d/visa_mobile_3d.htm>

[15] Weimerskirch, A., Parr C., and Shantz, S.C., Proceedings of the 6th Australian Conf. on Information Security and Privacy, July 11-13, 2001.

[16] Woodbury, A.D., Bailey, D.V., and Paar, C., "Elliptic Curve Cryptography on Smart Card without Coprocessors", Proc. of the 4th Smart Card Research and Advanced Applications Conf., September 20-22, pp1-20, 2000.

[17] Xydis, T.G., "Security Comparison: Bluetooth Communications vs. 802.11", 2002, <http://www.ccss.isi.edu/papers/xydis_bluetooth.pdf>

[18] Yeun, Chan Y. and Farnham, Tim, "Secure M-Commerce with WPKI", 2001, <http://www.iris.re.kr/iwap01/program/download/g07_paper.pdf>

Simulating Collaborative Mobile Services – An Approach to Evaluate a New Location Enabling Service

Peter Dornbusch, Florian Fuchs, Martin Huber

Center for Digital Technology and Management (CDTM), Technical University Munich,
Arcisstr. 21, 80333 Munich, Germany
{Peter.Dornbusch, Florian.Fuchs, Martin.Huber}@cdtm.de

ABSTRACT

Although enthusiasm for the “killer-application” in the mobile sector has decreased, location based services still seem to be valuable mobile services. Today’s location based services are mostly forced to use the proprietary location information provided by mobile network operators. The approach that is discussed here depicts a way of locating a mobile client independently from certain network operators and in heterogeneous networks. The Location Trader system collects and provides location information for users and providers of location based services. The accumulated interaction of each user of the user community enables to generate reliable location information in the Location Trader database. This paper prepares the groundwork for modeling the core principles of unintended value co-production with the Location Trader System. We intend to focus our analysis on the Location Trader System resp. Services, as the phenomenon of value co-production is widespread because of low coordination costs for integration and transactions enabled by immateriality and potential automation of business and transaction processes. The focus of this paper is the simulation of the user integration based on the theoretical basis.

Keywords: Location Aware, User Integration, Collaborative, Mobile Service

1. INTRODUCTION

Location-based services are considered by many as the next generation of mobile services with a huge potential: Allied Business Intelligence, for example, predicts that the market for location-based services will grow from one billion US dollars in 2000 to about 40 billion dollars in 2006.

Currently the development of such services, however, is severely impeded by the lack of an open, network-crossing positioning platform: Only splintered solutions are available for particular networks and the various operators of these networks are eager to defend their positioning monopoly.

In a novel approach the Location Trader system aims to overcome this situation and to establish a generic positioning platform that fosters the development of innovative and diverse location-based services.

The Location Trader approach is based on the integration of service users in the creation of the positioning service. It has first been introduced by Dornbusch and Huber in [1].

The Location Trader system constitutes a virtual link between service users, service providers and developers in a complex network constellation of these economic actors. This property together with the fact that the service is build up in cooperation with users poses several challenges for the successful launch and operation of such a system.

We identify two key properties of the Location Trader concept:

- Quality of service and therefore the value of the positioning platform increases with the number of people involved, and
- The system relies completely on volunteer contributions by its users.

In this paper we use the Location Trader system as a case study for analyzing unintended value co-production both qualitatively and quantitatively. We first introduce the relevant economic literature in section 2. In section 3 we apply these theories to the Location Trader concept. Section 4 models the concept quantitatively in a computer simulation and analyzes the results. The findings are summarized with a conclusion in section 5.

2. VALUE CO-PRODUCTION VIA USER INTEGRATION

Value co-production is different from the traditional view of value creation as a value chain [2]. Value co-production adds a new dynamic to the producer/customer relationship by engaging customers directly in the production or distribution of value [3][6]. Some researchers have given an even broader definition as “value co-creation by two or more actors, with and for each other, with and for yet other actors”[4].

Value co-production exhibits the following characteristics: Value is created more and more synchronously, it is created more interactively and it is created involving more actors. This has been enabled by recent technical breakthroughs and social innovations.

From a more implementation-oriented perspective, value co-production requires motivating customers as well as monitoring and managing the process of co-production. There are three generic ways of integrating the customer in the creation of value [5]:

- *Innovation* which describes an intended, innovative and creative contribution like generating new service ideas
- *Substitution* where a certain activity that used to be carried out by a service provider is now carried out by the customer, representing an intended but not innovative way of value creation
- *Variation* which describes a non-intended, non-creative and non-innovative contribution. “The value of a single customer activity is so small that the customer is not able to exploit his contribution and only the aggregation of many single parts by the service provider allows realizing the value”[5]

Some research has also been done on the potential of co-created value over time. For this Huber and Sandner have investigated *variation* as the special case of value co-creation where actors unintendedly and probably even unconsciously participate in value production [5]. This *non-intended value co-production* is made possible by the automation of certain business processes and the immateriality of the generated value, namely information.

A service based on variation usually exhibits feedback effects as contributions by users are aggregated over time and used to refine the service. Considering this feedback loop for the analysis of the service potential over time suggests the following three-phase model:

- **Build-up of service:** The initial service potential is low and will be build up over time through user input
- **Stable system:** User input and potential loss are balanced so that the system remains stable without the input from the service provider
- **Skimming of potential:** The user-created potential can be skimmed by the service provider

3. QUALITATIVE ANALYSIS OF VALUE CO-PRODUCTION WITH THE LOCATION TRADER

The economic ecosystem of the Location Trader concept involves two economic actors around the Location Trader positioning platform (see Figure 1): *Service providers*, who base their location-based services on the Location Trader positioning platform; *service users*, who access these services and contribute their current locations if necessary.

By design the Location Trader system relies on contributions by service users: They are the ones disclosing their current location when their current position –in the case of cellular networks the current

network cell –is not known to the system yet. This way they contribute to building up the location data pool and –together with all others service user –co-create value for all involved parties by improving the quality of the positioning service.

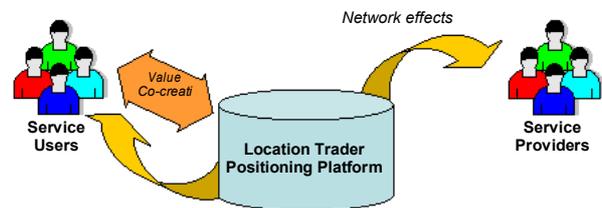


Figure 1: Economic ecosystem of the Location Trader approach

Co-creation in the case of Location Trader is enabled by cellular communications technology and the Internet in connection with the fact that the created value is immaterial information which can be transmitted and processed very cost-efficiently using these technologies.

The Location Trader approach exhibits the properties of value co-production, especially through value being created synchronously and involving all service users at the same time. This suggests modeling the value creation process as a value constellation rather than as a value chain where management has to focus on managing the interactions within this constellation.

Analyzing this value creation constellation from an implementation point of view suggests considering the following *risks*: Privacy concerns of service users may prevent them from contributing to the system. When they feed the data pool with location information they effectively disclose their current location. As they are requesting a location-based service this is actually required and will be plausible to them. However, as the system stores this information in the data pool, service users may be concerned that this information is used to track or trace back their movements. It needs to be credibly assured that the information they contribute is stored anonymously and therefore cannot be associated with their identity at any point of time.

Legal risks play a minor role in the co-creation process of Location Trader as the contribution process is well defined and the contributed information is always mere location information. However, it has to be ensured that intellectual property issues are considered.

The main *challenges* for co-creation in the context of Location Trader are goal divergence, effort and equity returns.

Both service users (as the value co-creators) and the positioning platform have different goals: Service users want to receive information or services based on their current location; the positioning platform wants users to access services and disclose their current location in

order to feed the location data pool. Nevertheless both goals can be combined and jointly satisfied with the Location Trader co-creation process where contributions are only required if nobody else has made a contribution yet.

Crucial for motivating users to contribute is to minimize the effort of contribution: Contribution in the case of the Location Trader co-creation process is a very structured and invariable task that can be highly automated. As most users are naturally used to disclosing their current location when requesting a location-based service they do not even perceive an additional effort. In fact, each user's average effort is reduced as contributions do not have to be made every time, but will automatically be reused in the future. There is a greater effort on the side of the positioning service to enable the co-creation of value through service users, but this effort has to be accepted as it constitutes the core of the whole concept. Interrelated with these two challenges is the issue of equity of returns. The return to each service user for the disclosure of their current location is the provision with a location-based service. It has to be ensured that the value of these services is perceived equitable to effort by each service user.

Applying *best practices* of value co-creation to the Location Trader concept yields the following insights: The objective of co-creation in the case of Location Trader is clearly defined: Service users disclose their current location so that the positioning platform can associate this location with the user's current positioning in a cellular network and can reuse this information in the future for positioning services. So co-creation takes place at the production stage. It also takes place at the maintenance stage once there is sufficient data available and further contribution are just used to refined the data pool and keep it up to date.

This implies that within the Location Trader ecosystem only service users can be selected as co-creators. It may make sense to further segment this group in order to select those of them who are most likely to contribute frequently.

The different rights of the involved parties need to be considered: Service users have the right of privacy and that their contributions are not associated with their identities. The positioning platform has the right of storing and using the contributed information.

Expectations of both involved parties need to be explicitly managed: Service users should expect that they are provided with a location-based service, but may sometimes have to disclose their current location. This will be the case more and more rarely in the future. Vice versa the positioning platform can expect every user to disclose the current location when it is required.

The channel that the positioning service can use to monitor and control user contributions is the list of incoming locations. They should be checked for plausibility before they are added to the core data pool.

Service users need to be enabled to contribute by providing them with the required Location Trader Client Applications and by making the contribution process as easy and self-explanatory as possible.

The users' incentives to contribute need to be explicitly communicated and managed. There is the immediate incentive of being provided with a location-based service that otherwise would not be available. However, there are also more indirect incentives such as the idea of the Location Trader platform as an independent, open source-like data pool where contributors can identify themselves with this vision. A competitive and signaling incentive can be created by allowing the association of a particular network cell with the nickname of the user who first "discovered" and contributed the cell to the data pool. Other incentives can be thought of.

An important aspect of value co-creation in the case of Location Trader is that it takes place automatically and not necessarily intentionally: The basic co-creation process is automated and users need to be involved only if their current location is not known to the system yet. Only then they become aware that they need to submit information in order to be provided with a location-based service. However, they are not necessarily aware that this information will also be reused at a later stage and therefore they are not aware that they are actually co-creating value. This kind of integration is denoted as *variation* as it is non-intended, non-creative and non-innovative.

Theory suggests that the potential of a service with these properties will follow a pattern of three phases over time: A buildup phase where initial potential is low and increased over time through user integration. A stability phase where the service remains stable just based on user contribution. Finally a skimming phase in which the service provider can skim the co-created value.

This section has applied the theoretical basis on value co-creation as presented in the previous section to the Location Trader system. It has pointed out risks, challenges and implications based on best-practices. Finally, it presented a three-phase model of the potential of the service over time.

4. QUANTITATIVE COMPUTER SIMULATION OF VALUE CO-PRODUCTION WITH THE LOCATION TRADER

In addition to the theoretical and qualitative analysis of the Location Trader system in the previous section, this section presents a more quantitative analysis based on a

computer simulation as a second approach to understanding the system's mechanics.

The simulation models the service usage behavior of a certain number of Location Trader users within a certain area and over a certain period of time. It takes into account several factors such as contribution of locations, individual satisfaction and completeness of the data pool. By varying initial settings and comparing the effects on the resulting simulations, the importance of particular factors is determined.

It is important to bear in mind that all conclusions are based on an idealized model of the real world and cannot be taken for granted. Nevertheless they can help to identify general interdependencies and trends. Furthermore the simulation should serve as an interactive model for playing around and experimenting.

4.1 Model

The model of the Location Trader system assumes a certain number of users in a certain quadratic area. This area is partitioned in equal-sized squares where each square represents an idealized network cell in a cellular network. Every user is unambiguously located within one of these network cells, however, as in real life one cell can accommodate several users.

Each user in the model has the properties requests, contributions, satisfaction and frustrations. *Requests* counts for each user how frequently the user requested a services. *Contributions* represents the number of locations that the user contributed to the data pool. *Satisfaction* measures the satisfaction of each user by incrementing the current value each time the user requested a service and was automatically provided with the service without the need of contribution. *Frustrations*, on the contrary, counts the situations when a user requested a service and had to provide the current location manually.

Each cell has the property *locations* which counts how many locations have been mapped to the cell's cell ID. If a cell is mapped to at least one location the model assumes that a service request in this cell can be satisfied.

When the simulation is started, each user acts autonomously, but everybody follows the same rules: First, the user roams around, which may or may not result in changing the network cell. Second, the user may or may not request a service which will either satisfy the user, if at least one location has already been mapped to the current cell, or will frustrate the user and cause the user to contribute the current location to the data pool. If the discrepancy between *satisfaction* (successful request of services) and *frustrations* (need to manually contribute the current location) grows too far, a user may be disappointed and stop using the system.

On the other hand, if *satisfaction* exceeds *frustrations* by a certain value, the user starts using the system more frequently and may even recruit new users to the system. Finally, after all users are done, a random network cell may be removed from the data pool to simulate changes in the network infrastructure.

The environment and exact behavior of the users is determined by several initial settings and also randomized between certain boundaries.

The cellular network environment is modeled by the radius of each cell and the number of hours after which the data pool is degenerated to simulate a change in the network infrastructure. In addition, the initial number of users, who will be randomly distributed over the network, can be set.

The behavior of the users is mainly influenced by the average number of times per week that they request a service (*usage*) and by their *mobility* that influences how frequently they move to another cell. Of course, this depends also on the radius of the cells.

In addition, it can be set how many more frustrations than satisfactions a user can stand before abandoning the system (*discouragement threshold*). The *recruitment threshold* determines how many more satisfactions than frustrations a user has to experience in order to start recruiting new users. The probability of a successful recruitment can be set, too.

Finally, it can be varied how frequently the systems asks the user to contribute the current location although the data pool already contains a location for the current cell (*redundant location request*). This is to maintain the integrity of the data pool and also to increase positioning quality.

4.2 Implementation

The previously described model was implemented using the software StarLogo 2.0¹ by the Media Lab at the Massachusetts Institute of Technology (see Figure 2 for a screenshot). StarLogo is a programmable modeling environment designed to help modeling and exploring the workings of decentralized systems, such as bird flocks, traffic jams, and market economies. It is available free of charge for download.

StarLogo is an extension of the Logo programming language and adopts Logo's terminology of autonomous *turtles* living on *patches* and an *observer* observing and controlling what happens.

In the implementation turtles represent service users and patches represent networks cells. The observer controls the overall simulation flow and provides a global clock.

¹ StarLogo website: <http://education.mit.edu/starlogo/>

The simulation is partitioned in discrete steps where each step represents one hour of simulated time. In each step each turtle completes one sequence of roaming and then probably requesting a service.

The parameters described as the initial settings for the model are implemented as *Sliders* on the simulation area and can be adjusted by the user. Service users/turtles and network cells/patches are visualized on the so called *Graphics Canvas*. The color of the patches is scaled between black and light green where the level of green indicates the number of locations mapped to a patch. The color of the turtles is scaled between dark red and white where the brightness indicates the level of satisfaction of a turtle.

So called *Monitors* display indicative figures such as the total number of requests, users and contributed locations. Furthermore the completeness of the data pool is measured as the percentage of cells that are mapped to at least one location. A fictive value of the current data pool is calculated by valuing the first location for each patch with a value of 5, the second location with a value of 3, the third with a value of 2 and the fourth with a value of 1. Finally, overall and average value for each user's contributions, satisfaction and frustrations are indicated.

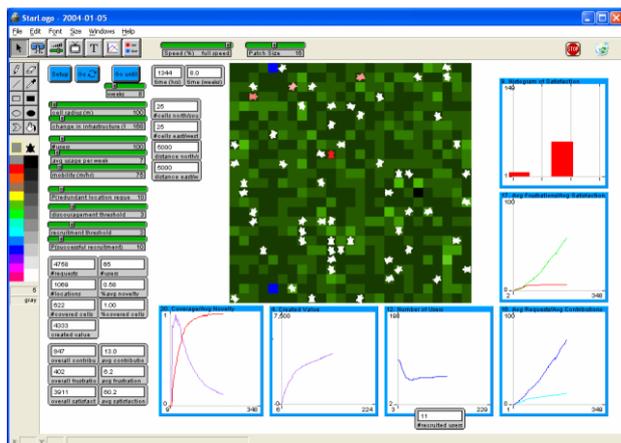


Figure 2: Computer simulation of the Location Trader concept in StarLogo

The temporal development of some of these figures is visualized in small *Graphs*. (Note that these *Graphs* can only use running time as their axis in x direction in contrast to simulated time.)

Three *Buttons* allow controlling the simulation: Setup resets and initializes the simulation with the current parameters. Go starts the simulations and runs it until the button is pressed again. Go until starts the simulation and runs until the simulated time has reached the predefined number of weeks.

4.3 Simulations

As a structured approach to simulation, two scenarios

and a set of basic initial settings have been chosen. They serve as a basis for comparison with simulations where selected parameters have been varied.

The two scenarios are an urban and a rural scenario. The urban scenario tries to model an urban environment where there are smaller and more network cells in a certain area. Based on reported average cell sizes the cell radius has been set to 100 meters with 25 times 25 cells. The resulting area of 5 times 5 kilometers approximates the area covered by the city of Munich.

The rural scenario is modeled after a rural environment where the radius of a cell can be several kilometers. Here a cell radius of 3000 meters has been chosen and the covered area is 15 times 15 kilometers.

The following basic initial settings have been chosen: The number of users has been set to 100. This seems to be a reasonable number that should be possible to reach with a concentrated effort at the university, the CDTM community and word-of-mouth propaganda. The average usage of Location Trader-based services has been set to 7 times a week which corresponds to 1 usage per day. This is a reasonable number, especially when motivation and curiosity are still high in the beginning. The average mobility of users has been set to 75 meters per hour which means that users move at most 75 meters per hour on average. This setting takes into account that most people spend several hours a day in the same building, where only few cell changes occur, but sometimes cover longer distances, for example during lunch break or in the morning and in the evening. Finally, the system has been set to require the user to contribute the current location redundantly in 10 percent of the cases where a location has already been mapped to the cell. It is assumed that a user accepts three cases more in which the current location has to be provided manually than cases in which a service is provided automatically. And it has been estimated that a user has to have been satisfied by the system three times more than being disappointed in order to start recruiting new users.

Based on these basic initial settings several simulations have been run. A comparison of the results of simulations with the same initial settings has shown that outcomes are converging and are therefore representative for the selected initial settings.

4.4 Evaluation

In 20 different simulation setups five basic initial settings, as explained in the previous section, have been varied along a single dimension each time in order to single out and identify their effects on the simulation outcome. Both urban and the rural scenario were considered. The simulation was run for a simulation time period of 8 weeks to allow a comprehensive overview of the launch period.

The five settings that were varied are: the initial number of users, frequency of usage, mobility of the users, the users' discouragement threshold and the redundancy of contribution (the percentage of cases in which a user is required to provide the current location although at least on location information is already available). All these settings were both halved and doubled one after the other and the effects on the simulation outcome were examined.

The examination was based on the value of nine indicative figures at the end of the simulation: the completeness of the data pool, the total created value based on the valuation as described in 0, the number of users, the total numbers of requests and contributions and the users' average number of requests, average number of contributions, average satisfaction and average frustration.

4.5 Urban Scenario

The outcomes for the urban scenario suggest the following interdependencies (see also Table 1):

Both the initial number of users and the frequency of usage seem to have the greatest effect on the completeness of the data pool. While an almost complete coverage is achieved in almost all cases after at least 8 weeks, these two factors speed up the process. They also increase the total created value and the overall number of requests and contributions. The initial number of users has an even higher impact on the latter two, whereas the frequency of usage increases the average number of requests per user. This is understandable as both factors have a direct effect on the number of locations contributed to the system and therefore have a direct impact on the quality of service and the satisfaction of users. Frequent usage further pushes average satisfaction which can be explained with the more frequent satisfaction that users experience. On the other hand, a greater initial number of users helps to extremely decrease average frustration as the disappointment of having to provide the current location manually is spread across more people.

In fact, the initial number of users is mission critical as in all simulations for the urban scenario a great loss of users occurs in the beginning. If the initial number of users is too low (less than 75 in the simulation) the initial great loss may even be unrecoverable and the system may fail completely. (The user loss is due to the frequent frustrations that user experience with a not-yet-complete data pool in the beginning.) On the other hand, the higher the initial number of users is, the more users the system will have in the future. In fact, this seems to be a super-proportional relation.

Surprisingly, user mobility as an indicator for the spatial distribution of service usage has very little effect on any

of the indicative figures. It seems that a high number of users already provides sufficient spatial distribution.

The discouragement threshold proves to be another mission critical factor. Simulations show that users must accept at least two frustrations more than satisfactions before abandoning the system to keep the system running. Otherwise the number of users drops dramatically and the system fails. In the simulation a discouragement threshold of three is sufficient. A higher threshold has only a moderate positive effect on the number of users and requests.

Requiring the user to redundantly provide the current location has an interesting effect on several factors. Of course, it increases the average number of contributions per user and maintains the quality of the data pool. On the other hand it decreases average satisfaction and therefore the number of users. But this is only true for redundancy rates of more than 20% in the simulation. Below this value these negative effects are negligible, but improvements in completeness and quality of the data pool are still gained. A moderate rate of redundancy does therefore make sense.

In conclusion in the urban scenario the initial number of users and frequency of usage are the most effective levers on the number of service requests and the future number of users. The initial number of users is mission critical, but at the same time most effective for increasing the number of users. There needs to be a certain level of resistance of users against frustrations in the beginning. Redundant manual contributions seem to be acceptable to users within certain boundaries.

4.6 Comparison to Rural Scenario

The basic observations for the urban scenario proved to be true for the rural scenario, too. However, there are some important differences that will be pointed out in the following.

First of all, simulations for the rural scenario never failed: Initial number of users and their discouragement threshold are not as mission critical as in the urban environment. In fact, in the rural scenario there is basically no loss of users in the beginning, which is very much in contrast to the urban scenario. On the other hand, a high initial number of users does not have a significant impact on the number of users in the future. This can be explained with the fact that users in the rural scenario only infrequently change network cells due to the size of the cells. This prevents them from frequent frustrations by the system, but at the same time does not make them enthusiastic enough to attract new users. In addition, it generally slows down the build-up and completion of the data pool.

Table 1: Simulation outcomes for the urban scenario

Effect of higher ... on ...	Number of users	Frequency of usage	Mobility of users	Discour. Thresh.	Redund. of contrib.
Completeness of data pool	proportional	super-prop.	insignificant	insignificant	insignificant
Created value	sub-prop.	sub-prop.	insignificant	insignificant	insignificant
Number of users	super-prop.	sub-prop.	insignificant	sub-prop.	insignificant
Total number requests	super-prop.	proportional	insignificant	sub-prop.	insignificant
Total contributions	super-prop.	proportional	insignificant	sub-prop.	insignificant
Average requests	insignificant	proportional	insignificant	insignificant	insignificant
Average contributions	insignificant	sub-prop.	insignificant	insignificant	proportional
Average satisfaction	insignificant	super-prop.	insignificant	insignificant	insignificant
Average frustration	invers Prop.	insignificant	insignificant	insignificant	insignificant

As users usually stay close to their home cell, the system does not require a comprehensive data pool to provide them with satisfying services. This decreases the importance of the discouragement threshold in the rural scenario. The negative effects of requiring redundant contribution are also reduced. However, due to the size of network cells in the rural scenario, positioning accuracy in general is dramatically lower which affects the overall value of services to users.

All in all, general trends and interdependencies are the same for both the urban and rural scenario. However, the rural scenario is less prone to failure, but at the same time provides lower quality of service due to lower positioning accuracy.

5. CONCLUSION

Value co-creation takes place between service users and the positioning platform. The major risk involved with this is that privacy concerns of service users may prevent them from contributing to the system. It has therefore to be credibly assured that the location information cannot be associated with their identity at any point of time. The Location Trader approach naturally ensures that the goals of both parties are satisfied in the co-creation process: the positioning platform improves its data pool and service users are provided with a location-based service. Requiring a contribution only if necessary further reduces the effort of co-creation. In order to ensure equity of returns, the quality of the offered services has to be guaranteed. For encouraging service users to contribute their expectations have to be managed and they have to understand why they sometimes have to disclose their location. A major incentive will be to fascinate them for and make them identify themselves with this

independent and open source-like approach. Still, contributions have to be checked for intentionally corrupt information. Finally, as the co-creation process is nearly completely automated, some users may not even perceive it as such.

For quantitative analysis the service usage behavior of a certain number of Location Trader users within a certain area and over a certain period of time was modeled in an idealized computer simulation.

The model takes into account environmental factors such as network cell size and initial number of users as well as individual factors such as frequency of usage, number of requests, number of contributions, each user's satisfaction level, frustration level and tolerance towards frustration before abandoning the system.

By systematically varying certain initial settings, their effect on the outcome of the simulation can be observed. Each simulation was run for both an urban and a rural scenario.

The basic observations are the same for both scenarios: Initial number of users and frequency of usage have the greatest effect on build-up speed and the future number of users. Users have to tolerate a certain number of frustrations in the beginning when required to manually provide their location. However, a certain level of redundancy in requiring contribution is acceptable. In contrast to the rural scenario, the urban scenario may fail if the initial number of users or their tolerance towards frustration is too low. However, positioning accuracy is significantly lower in the rural scenario due to the greater size of network cells.

REFERENCES

- [1] Huber, Martin / Dietl, Tobias / Kammerl, Julius / Dornbusch, Peter: *Collecting and Providing Location Information: The Location Trader*, in: Proceedings of 8th International Workshop on Mobile Multimedia Communications, pp305-310, 2003.
- [2] Porter, Michael: *Competitive Advantage*, New York (Free Press), 1985.
- [3] Kambil, Ajit / Friesen, G. Bruce / Sundaram, Arul: *Co-creation: A new source of value*, in: Outlook 1999, Number 2, 1999.
- [4] Ramirez, Rafael: *Value Co-Production: Intellectual Origins and Implications for Practice and Research*, in: Strategic Management Journal, Volume 20, 1999, pp 49-65.
- [5] Huber, Martin / Sandner, Uwe: *Constellations for Non-Intended Value Co-Production by Mobile Users*, CDTM Working Paper, Munich, 2003.
- [6] Friesen, G. Bruce: *Co-creation. When I and I make 11*, in: Consulting to Management, Volume 12, Issue 1, March 2001.

Supporting Salespersons CRM Efforts through Location Based Mobile Support Systems

Chihab BenMoussa

Institute for Advanced Management System Research, Åbo Akademi University, Finland
cbenmous@abo.fi

ABSTRACT

The paper aims at assessing how can location based mobile support systems support salespersons' CRM efforts when they are operating within a highly mobile work environment. The paper is structured as follows. In the first section the paper conceptualises key properties of location based mobile support systems. It then introduces the dual role of salespeople in CRM through a brief literature review. A fourth section suggests potential mobile location services and applications that can support salespersons in performing effectively their everyday CRM tasks and links such applications to the determinant of salespersons' performance. The papers concludes with some remarks and suggests some areas of further research

KeyWords: Mobile location technologies, Customer relationship management, Salespersons tasks, salespersons performance, Personalisation.

1. INTRODUCTION

Effective customer relationship management (CRM) has emerged as a strategic imperative for companies in virtually every business. The rapid rates at which innovations can be made and copied have made product advantages brief (Shoemaker, 2001). However, as Day (2001) observed, because nurturing customer loyalty is usually resource intensive, long term and difficult to manage, loyalty relations can provide a durable competitive advantage.

Today many companies are trying to move closer to their customers by capturing the available wealth of internal and external data to better understand their customer needs, preferences and profitability and then leverage that knowledge in every customer contact. Recent studies show that the movement to CRM is gaining momentum. One study by IDC Group forecasts that the worldwide market for CRM products and services would top US \$ 125 billion while a survey of more than 1500 companies by the Data Warehousing Institute in 2000 found that 91 percent either have or plan to deploy CRM solution in the near future (Goodhue et al. 2002). The impetus for this interest in CRM came from Reichheld, who demonstrated a dramatic increase in profits from small increases in customer retention rates. His study showed that as little as a 5 percent increase in retention had impacts as high as 95 percent in the net present value delivered by customers.

Salespersons play a dual role in CRM. On one hand, forging close relationships between firms begin with the initial contact by a salesperson (Levitt, 1993) who serves as a boundary spanner between the selling firm and the customer. For some customers the salesperson is virtually synonymous with the firm (Crosby et al., 1990; Czeipeil, 1990). On the other hand salespersons possess

valuable source of market information that is frequently needed in strategic planning and CRM programs. However crucial marketing information still remains locked in the heads of salespersons in the form of tacit knowledge. Converting tacit dispersed information possessed by salespersons into explicit knowledge that can be used to better satisfying customer is still a challenge for a number of companies (Choo 1996).

The advent of access to services through mobile and wireless devices has resulted in a fast growing of a number of mobile applications and services. Mobile (or wireless) applications, despite being different in their nature, they share a common characteristic that distinguishes them from their wire-line counterpart: They put the user at the centre of information and communication through the provision of location specific information, personalization, immediacy, and service availability (Durlacher, 2001). These characteristics would enable the development of innovative mobile applications to support firms' salespersons in developing and nurturing committed customer relationships. How and which mobile applications can support these frontline ambassadors in their CRM related tasks are key questions facing a number of stakeholders including sale managers today.

The purpose of this paper is to assess how can location based mobile support systems support salespersons' CRM efforts when they are operating within a highly mobile work environment.

The remainder of the paper is structured as follows. Next the paper conceptualises key properties of location based mobile support systems. It then introduces the dual role of salespeople in CRM through a brief literature review. A fourth section suggests potential mobile location services and applications that can support salespersons in performing effectively their everyday CRM tasks and

links such applications to the determinant of salespersons' performance. The papers concludes with some remarks and suggests some areas of further research.

2. KEY PROPERTIES OF LOCATION BASED MOBILE SUPPORT SYSTEMS

2.1 Personalisation

Personalisation in information systems is about adapting content and services based upon the user's interests, preferences, and behaviour, allowing further optimisation, which results in an optimal experience for the user, targeting the goal of the user (Setten, 2003). One key enabler of providing personalised services to users is ensuring the adaptivity of the system. System adaptivity is the process in which a system adjusts itself based upon some form of model of the user's needs. Location awareness can be a key feature in updating the mobile user's model based on the system's awareness of the user's current spatial context. Information about the user's current geographical position when combined with other elements such as the user's daily schedule of activities, or her predefined preferences in terms of content or interests would enable the service provider to proactively deliver a meaningful mobile support which is based on a continuous update of the user's changing spatial environment. For instance, within the sales force work environment, the service provider can deliver, based on the awareness of the salesperson's geographical position and her predefined preferences about the information she wishes to receive, a personalised support that has the potential to add value within the spatial context where the salesperson is located. Also the support can be delivered in the form that matches the requirements of the salesperson work context (i.e. face to face meeting with a client, in the train or in restaurant). For instance, the service provider can push an alert to the salesperson's smart phone in the form of text message if the salesperson is in face-to-face meeting and thus enabling him/her to read the alert and potentially exploit it during his/her interaction with the client.

2.2 Relevance

Knowledge of the user's spatial position can be a key element in providing relevant support to him/ her. Awareness of the user's geographical position can provide a good filtering mechanism in terms of determining how relevant is the intended mobile support given the user's current geographical location. This has the potential of reducing the functional deficiency associated with information overload, where the amount of information the individual client side user encounters extends her cognitive processing capacity (Ljungberg and Sorensen 2000). For instance, combining the knowledge of the salesperson's current geographical location with her calendar of daily activities can provide the service provider with insights that the salesperson is

about to meet the client X whose address matches the geographical position where she is located. Then the service provider can push to the sales person relevant information that fit him /her predefined preferences about that client that he/she could use during the sales interaction.

2.3 Convenience

The ability of the service provider to know the geographical position of the mobile user can make it faster and simpler for such user to be provided with the targeted support she needs. Indeed it will suffice the mobile user to notify its request to the service provider to get the service he /she is seeking as the service provider can know where the user is located. Indeed, the user's request can be automatically completed by the current location (i.e. if the user asks about the restaurant nearby, the system can infer from the user's geographic position what nearby means and select such restaurants that are in the area (BenMoussa, 2004).

2.4 Timeliness

Another key characteristic of location-based mobile support is the timeliness of the support that service providers can provide users with, based on the knowledge of their locations. The timeliness of location-based mobile support refers to the ability of the service provider to provide the user with support at the moment of value. The moment of value can be defined as "the moment when I a service provider can do something for you where you are and regardless of where I am or what time it is"(Keen, 2001).

2.5 Proactiveness

Another key property of location based mobile support system is the ability of the service provider to proactively enable the user to react to changing trends in the environment. Based on the user's situational context and changing environment, the service provider delivers content without receiving a request from him/her. Providing a proactive support to users can be of great importance in terms of saving time (i.e. alerts about traffic jam), tapping an opportunity (i.e. notifications about nearby tradeshows) or avoiding relationship-damaging problems (i.e. getting in time to a meeting with a new client thanks to the road instruction delivered by the service provider).

3. THE DUAL ROLE OF SALESPERSONS IN CRM

Evidence indicates that salespeople play a central role in the evolution of quality business relationships (Crosby et al, 1990;Dwyer et al., 1987). According to Kotler (1994), personal selling is evolving towards the salesperson as "relational manager"who is able to build strong ties with important customers. Further, Swan and Nolan (1985) contend that firms are seeking out relationships with

their customers and that salespeople are important in helping to build these relationships. One reason for the critical nature of a salesperson's role in developing quality relationship between their firm and the customer lies in the salesperson's ability to develop strong lines of communication with the customer. In addition to their role as relationship-builders, salespeople serve as CRM-enablers through the wealth of market information that they possess and which can serve as a basis in developing effective CRM programs. The following discusses such a dual role of salespersons in CRM.

3.1 The Salesperson as Relationship-Builder

In the marketplace, salespersons often perform a vital boundary-spanning role that links their firm to customers. For some customers the salesperson is virtually synonymous with the firm (Crosby et al., 1990; Czeipeil, 1990). Research on buyer-seller relationships suggests that a firm's salespeople are uniquely positioned to address several issues related to quality relationship building. Through their interaction with buyers, salespeople can increase the customer's confidence in the supplier. This helps reduce uncertainty and increase trust (Moorman et al, 1992). The relationship-builder role of salespersons is critical throughout the relationship evolution process.

3.2 The Salesperson as CRM-Enabler

The salesperson's role as CRM-enabler stems from the knowledge he/she possesses about the marketplace and which can serve as a valuable input in feeding customer information system for CRM. Marketing scholars have indeed advocated that business salespeople be incorporated into firm's formal marketing information systems for almost 40 years (Webster 1965). For instance klompaker (1980) stated, "Salespeople are used because they possess a great deal of information about the market". To leave this rich source of information untapped would be fool-hardy"(Evans et al, 1985).

4. LOCATION BASED MOBILE APPLICATIONS TO SUPPORT SALESPERSONS CRM EFFORTS

After discussing the key properties of location based mobile support and the dual role of salespersons in CRM, the paper now presents potential mobile location applications and services (MLS) to support salespersons tasks for better performance. The paper categorises such mobile location applications and services as the adaptive, the proactive, the locative and the disseminative. The paper shows how each category of such mobile location based applications can support salesperson CRM tasks and enhance their overall performance through its impact on a number of mediator variables that constitute salespersons' performance determinants. Indeed, several authors (Mooney et al. 1996; Huber, 1990; Davenport 1999) have proposed that in order to uncover the added value mechanisms and the impact of information

technology on productivity, studies should include intermediate benefits of information technology. For instance according to the theory of the effects of advanced information technology (Huber's 1990), the benefits in individual and organizational effectiveness occur" indirectly "through the positive impact the technology has on information and communication processes. Figure 1 depicts a model linking location-based mobile support system with salespersons' performance.

4.1 Adaptive MLS

Location based mobile adaptive applications are aimed at supporting situations where the salesperson is engaged in the interaction with her client and needs useful information about the client so that she can practice adaptive selling. Salespeople's adaptive selling is one of the main determinants of their performance (Weitz et al, 1986; Sipro and Weitz 1990; Sujun 1986). Indeed the process of selling requires that the salesperson matches the customer's needs with the available range of products. The process of adaptive selling involves two stages. In the first stage, the salesperson uses the available information to form impressions about the customer and in the next stage she formulates a sales strategy that maximizes the fit between the customer preferences and the company's offerings. Through the practice of adaptive selling, salespeople exploit the unique opportunities of personal selling. However it has been shown that adaptive selling can be improved by providing salespeople with the necessary market information and resources such that they can link insights from other sales situations to the customer contacts in which they are currently engaged (Weitz et al.1986). For instance, the market intelligence department can push, based on knowledge of the salesperson's location, an alert about a new discount policy released by the company's main competitor related to the product she is selling to her client. The salesperson can then read the alert and prepares suitable arguments should the client raises issues associated with new competitor's discount policy. Likewise, the salesperson can receive, based on her location, real time notifications about the business of the customer she is visiting and use such information in tailoring her proposal to the customer.

Location based adaptive applications can also support salespersons in altering their sales call schedule based on alerts about her customers changing orders and profitability as well as competitors' move, so that they can target their relationship efforts on the most profitable customers. For instance the market research department may push an alert to the salespersons that one of her customers, which is located in the nearby has just cancelled an important order. Such alerts would enable the salesperson to adjust her sales visits schedule so that she can devote more efforts to react to the potential account loss.

4.2 Proactive MLS

Proactive MLS enable salespersons to continuously search for market opportunities and experiment with potential responses to changing environment. (Venkatraman, 1989).

An indicative example of proactive location based mobile applications include notifications, based on the salesperson's geographical position, about sales leads which are located in the nearby. For instance, the market research department can push alerts about trade shows in the area where the salesperson is located. Similarly, telemarketing support centre can provide real time alerts to the salesperson about high -quality leads that are in the same area where the salesperson is located. Depending on the quality of the lead (that is, sales versus no sales) and the salesperson's sales calls schedule, the salesperson can accept or deny making face-to-face sales visit to the sales call identified by the telemarketing support centre. If the salesperson accepts to make the sales visit then additional information from the marketing research department can be pushed to his/her mobile terminal about the lead including a rebuttal to prepare him/her for question that the prospect may raise (BenMoussa 2004). Furthermore, proactive MLS may include location based intelligent alerts that would enable the salesperson to respond proactively to the competitors' alerts threatening her accounts.

Another indicative example of proactive MLS is notifications about traffic conditions (i.e. traffic congestion or a road blocking accident together with alternative routes.

Proactive MLS may include also notifications pushed by the different functions within the salesperson's organization. For instance the customer service department can alert the salesperson that the order of a customer located in the same geographical area where the salesperson is may experience a delay in delivery. This would enable the salesperson to contact the customer and proactively suggest alternative solutions. Similarly, the account department may send a notification to the salesperson that the client she is about to meet has not yet paid her invoice so that the salesperson can tackle the invoice payment within her interaction with the customer.

Proactive MLS would enhance the salesperson ability to focus her efforts on value adding activities such as maintaining quality relationship with her customer. In addition proactive MLS have the potential to enhance the salesperson customer orientation, which is a key determinant of salespersons performance. Support for the use of customer oriented selling is provided by the contention that customer orientation is an important characteristic of high performers (Kelly 1992, Bragg, 1986, Mackay 1988; Peterson 1988). For instance according to MacKay (1988) "best" salespeople are

genuinely interested in their customers and that sales representatives sell to people not to computers. Likewise Peterson (1988) contends that successful salespeople work to satisfy the needs of the customers

4.3 Locative MLS

Locative MLS are aimed at empowering the salesperson to deal customer requests and problems. The purpose is to shorten the time it takes to address customer problems and concerns. Indicative examples of locative MLS may include product tracking applications, customer support staff locator applications colleagues' locator, and entertaining areas for clients (BenMoussa, 2004). Product tracking applications consist of enabling salespersons, irrespective of their locations, to track the delivery status of products ordered by customers, either by connecting wirelessly to smart tags incorporated in the products or through receiving location based alerts about the order status of clients situated in the same geographical location where the salesperson is located. This would enhance their ability to answer rapidly, accurately and irrespective of their locations to customer order related inquiries, which their customer orientation. Furthermore, receiving location based alert about customers' orders status would enable salespersons to react to eventual shipment problems that may result in a delivery delay to customers.

Customer support staff locator applications would enable salespersons to locate the nearest customer support staff in order to address a client's problem. The salesperson can send a request to locate customer support staff, display their locations on a map and forward the client's request for service to the nearest field worker. Upon receiving confirmation to perform the service by the field worker, the salesperson can then be able to provide his/her client with accurate personnel arrival time. Dispatching the nearest customer support field employee has the potential to reduce the time needed in providing support to customers, which would enhance the customer's perception of the salesperson's empathy.

Colleagues locator applications would support the situations where the salesperson need to collaborate with her colleagues in order to deal with a complex sales problem. The salesperson can then identify the colleagues who are in the same geographic position as her and checks whether they are available for a meeting or a conference call.

4.4 Disseminative MLS

The disseminative MLS support the situation where the salesperson wants to alert her colleagues about a market opportunity and /or potential threats within the colleague's sales area. The salesperson can use the geographical position of her colleague as a filtering mechanism to assess the potential impact of the alert in terms of information overload. For instance the

salesperson can uncover a competitor's deficiency in the sales area of her colleague. The salesperson can then based on her colleague's geographical position and her calls agenda assess whether or not the insight she uncovers may help her colleague in practicing adaptive selling with the client she is visiting. Therefore the salesperson may either push directly the notification to her colleague or elect to submit the alert to the corporate customer database for subsequent forwarding to her colleague.

Disseminative MLS have the potential to exploit the wealth of information the salesperson possess as a result of their daily interactions in the market place. In addition, disseminative MLS provide salespersons with a tool that has the potential of enhancing collaboration among them despite their constant move. Furthermore, enabling salespersons to disseminate the market insight just after it occurs (i.e. after the sales interaction) would enhance their accuracy. The importance of accurate customer knowledge to salespersons performance is intuitive. If salespersons' perceptions are inaccurate, it can adversely affect customer relationship building at least in two ways. First inaccuracy in assessing customer requirements by the salesperson will negatively affect both relationships

building and corporate profitability particularly in industries where sales persons are given the latitude in tailor prices and services to individual customers. Second, the use of inaccurate salespersons' insight in the company's planning process including CRM programs will adversely affect the quality of these plans (Lambert et al.1990).

5. CONCLUSION

The paper explored the area of location based mobile support systems within the context of salespersons' CRM efforts, when they are operating within a mobile work setting. More specifically the paper discussed the unique role of salespersons as both an enabler and facilitator of CRM. The paper also suggests potential location based applications to support salespersons in their CRM efforts and linked them to the determinants of salespersons performance. Future research would involve issues associated with the acceptance by salespersons of location based mobile application and services so that they could integrate them into their everyday work life and take benefits from their full potential.

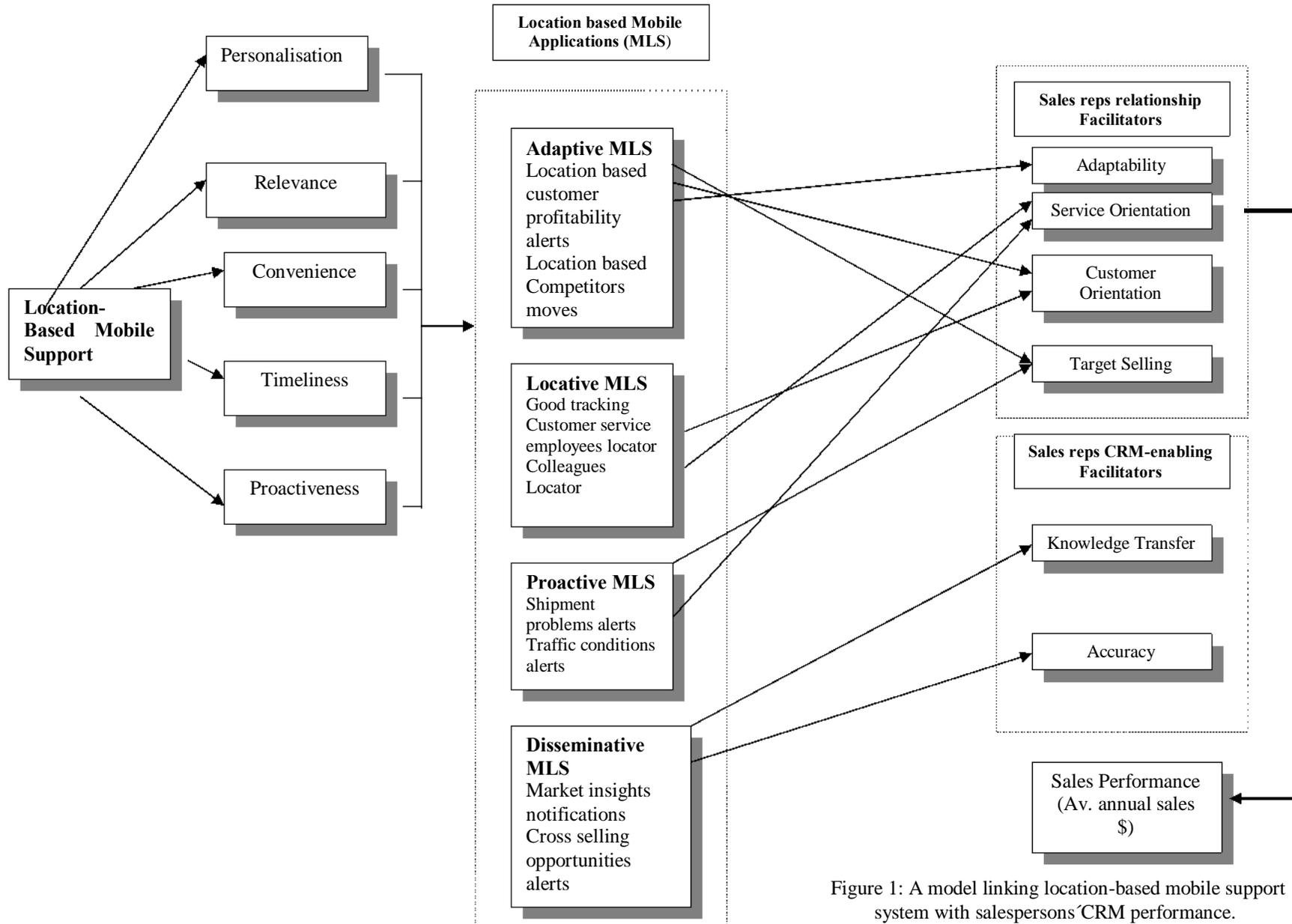


Figure 1: A model linking location-based mobile support system with salespersons' CRM performance.

REFERENCES

- Apicella.M, Mitchell.K, and Dugan, S (1999). "Customer Relationship Management: ramping up sales service", *Info World*, Vol 21No33, pp. 68-80
- BenMoussa. C., (2003) "Supporting salespersons through location based mobile applications and services", In the proceedings of IFIP world computer congress, Kluwer publisher, Toulouse, France, August 2004,
- Bitner, M., Booms, B. and Tetreault, M (1990). "The Service Encounter: Diagnosing Favourable and unfavourable incidents". *Journal of Marketing*, Vol.54, January.
- Bragg. A (1986). "Turning Salespeople into Partners", *Sales & Marketing Management*, Vol.137.
- Brown.S and Swartz (1989). "A Gap Analysis of Service Quality". *Journal of Marketing*, 53.
- Christopher. M, Payne.A and Ballantyne. D (1991). *Relationship Marketing*. Butterworth-Heinemann, Oxford.
- Crosby, L. A, Evans, K.R and Cowles, D. (1990), "Relationship quality in services selling: an Interpersonal influence approach" *Journal of Marketing*, vol. 54.
- Czeipiel, J.A. (1990), "Service encounters and service relationships: implications for research", *Journal of Marketing*, Vol.51.
- Cronin,J. and Taylor.S (1992). "Measuring Service Quality: A Reexamination and Extension". *Journal of Marketing*, Vol.56.
- Davenport C (1999). *Performance Measures for knowledge management*. Knowledge management handbook. Edited by Jay Liebowitz
- Day. G (2000). "Managing market relationships". *Journal of The Academy of Marketing Science* 28 (1), 24-30
- Eklund S. and Pessi. K. (2001). "Exploring Mobile eCommerce in Geographical Bound Retailing. Proceedings of 34TH Hawaii International conference on System Science.
- Girard, K (1998) "Stats not good for sales technology" *Computer world* 32, 14 (April 6), 29
- Goodhue. D, Wixom. B and Watson. H (2002). "Realizing Business Benefits Through CRM:Hitting the Right Target in the Right Way". *MIS Quarterly Executive*. Volume 1, Issue 2.
- Giaglis. G.and P. Kourouthanassis (2003). "Towards a Classification Framework for Mobile Location Services" *Idea Group publishing*
- Huber, G. (1990), *A Theory of the Effects of Advanced Information Technologies or Organizational*
- Kehoe.L (2002) "Long live e-business: software is finding a new role in helping companies to share information effectively". *Financial Times*, March p.14
- Kelly, S. (1992). "Developing Customer orientation among service employees. *Journal of the academy of Marketing Sciences*, Vol. 20 No 1, pp.27-36.
- Kotler, Philip (1994), *Marketing Management: Analysis, Planning, Implementation and Control*, 8 ed., Prentice Hall, NJ.
- Lambert, D., Marmorstein, H, and Sharma,A. (1990). "The accuracy of salesperson perception of their customers: a conceptual exam and empirical study". *Journal of personnel selling and sales management*. January,pp-18.
- Leijfer, SC (1999) "Bringing Business Intelligence to sale Force Automation" *American Salesman*, 44, 5 (May) 26-30
- Mooney, John G., Gurbaxani, Vijay and Kenneth L. Kraemer (1996), *A Process Oriented Framework for Assessing the Business Value of Information Technology*, The DATA BASE for Advances in Information Systems, Spring, 27, 2, 68-81.
- Raatikainen. K, Christensen. H, and Nakajima. T (2002). "Application Requirements for Middleware for Mobile and Pervasive Systems. *Mobile Computing and Communication Review*, Volume 6, Number 4.
- Rodina,E, Zeimpekis. V, Fouskas.K, (2003). *Remote Workforce Business Process Integration through*
- Jackson.C (1999). "Process to Product: Creating Tools for Knowledge Management"
- Keen P. et al, 1991, (2001). *The Freedom Economy*, Osborne/McGraw-Hill, USA.
- Laitinen. H, Ahonen. S and Kyriazakos. S (2001). "Cellular Locations Technologies
- MacKay, H. (1988). "Humanize your Selling Strategies". *Harvard Business Review*. Vol.66.
- Moorman, C., zaltman, G. and Deshpande R. (1992). "Relationships Between Providers and Users of Market
- Peterson. R (1988). "The Canapé approach to Personnel Selling". *Industrial Distribution*.
- Research: The Dynamics of Trust Within and Between Organizations", *Journal of Marketing Research*; Vol.29.
- Varshney. U, and Vetter R. (2001). "A framework for the Emerging Mobile Commerce Applications". *Proceedings of 34H Hawaii international Conference on System Science*.
- Reichheld.F (1996). "The Loyalty Effect". *Harvard Business School Press*.
- Rupnik.R, krisper.M (2003). "The Role of Mobile Applications in Information Systems". In the proceedings of The second International Conference on Mobile Business, June 2003,
- Saxe, R.and Weitz. B (1982). "The SOCO Scale: A Measure of the Customer Orientation of Salespeople". *Journal of Marketing Research*, Vol.19.
- Setten.Van (2003). "Personalised information Systems". *Telematica Institute*. Report reference: GigaCE/D2.3 and Duine/D1.
- Shoemaker.M (2001). "A Framework for Examining IT-Enabled Market Relationships". *Journal of Personnel Selling and Sales Management*. Volume XXI, number 2 (spring 2001, pages 177-185).
- Sorensen. C, Mathiassen. L and Masao.K (2002). "Mobile Services: Functional Diversity and Overload".
- Sujan, H., Weitz, B., Kumar, N. (1994), *Learning Orientation, Working Smart, and Effective Selling*, *Journal of Marketing*, 58, July, 39-52.
- Sujan, H., Weitz, B.A., and M. Sujan (1988), *Increasing Sales Productivity by Getting Salespeople to Winner*. R (2002). "A Framework for Customer Relationship Management". *California Management Review*. Volume 43. No 4.
- Weitz, B., Sujan, H, Sujan, M. (1986), *Knowledge, Motivation and Adaptive Behavior: A Framework for Improving Selling Effectiveness*, *Journal of Marketing*, 50, 174-191.

Technology Attribute-Diffusion Pattern Nexus: Evidence from the Early Experiences of 3G Mobile Communications

Nir Kshetri

University of North Carolina at Greensboro, NC 27402-6165, U.S.A
nbkshetr@uncg.edu

ABSTRACT

The diffusion patterns of 3G cellular technologies indicate that the reality is not matching the optimistic estimates of various consulting firms regarding the size of the network. What factors influence the diffusion of rapidly developing network goods? This paper draws upon innovation diffusion literature and the early experiences of 3G mobile technologies to examine the influence of supply-side factors on the 3G diffusion trajectory. We also propose a model that classifies various 3G related attributes in terms of their influence on present utility vs. expected future utility and the mechanism of influence: via 3G device vs. via 3G network. The central argument of this paper is that while the breadth of offers has been wider in 3G compared to lower generation cellular technologies, there has been digression rather than progress on some dimensions (e.g., bulky 3G handsets, very short battery lives, small geographical coverage, and lower network externality effects). The problems are compounded by the lack of strategic pricing and industry-regulator coordinations.

Keywords: 3G mobile communications, diffusion, supply-side, technology attributes, upgradability, utility of network goods

1. INTRODUCTION

The diffusion patterns of 3G cellular technologies indicate that the reality is not matching the optimistic estimates of various consulting firms. The early experience in Europe and Asia-Pacific indicates dismal consumer response to 3G mobile communications. For instance, Hutchison 3G (H3G), which launched 3G services in the UK and Italy in March 2003, had set a target of target of 1 million subscribers in each country by December 2003. The company achieved only 30% of its target in Italy and 21% in U.K. by that time (Table 1).

This paper draws upon innovation diffusion literature and the early experiences in Europe and Asia to explore the connection between attributes of a technology and its diffusion pattern. Although we focus on 3G, the issues examined in this paper have broad implications technology marketers.

2. SUPPLY SIDE OF TECHNOLOGY DIFFUSION: RELEVANT THEORIES

The existing literature indicates that supply side factors that influence diffusion of network goods include price, breadth and depth of technology products, complexity, positioning, government-supplier coordination and modularity and upgradability, degree of supply abundance and expected size of the network (Table 2). It should, however, be noted that these factors are not necessarily orthogonal. In the case of network goods such as 3G mobile communications, some of these

factors are more relevant to the device whereas others are more relevant to the network. These attributes can also have differential impact on the perception of present utility and expected future utility. Table 3 classifies the attributes in two dimensions: device specific vs. network specific and present utility vs. expected future utility.

The demand for a network good is a function of the expected size of the network (Katz and Shapiro 1994). Too small new market may result in the failure of a new technology (Lin 2003). The expected size of the network is a function of various factors including price of the technology product, quality, degree of compatibility with previous generations of technologies, and upgradability.

Suppliers can modify product characteristics and price structure to fit the needs of a market segment (Robertson and Gatignon 1986). The price of a technology influences consumers' adoption timing (Kamakura and Balasubramanian 1988) and the rate of adoption (Jain and Rao 1990) at the micro-level and for a network good, the demand and the market potential (Kalish and Lilien 1986; Katz and Shapiro 1994) at the macro-level. Since the network externality is positively related to the network size, the number of users has a higher significance than the intensity of use for network goods. For networks that require two part fees, a reduction of the fixed component is thus a more appropriate way to stimulate the growth of the network than a reduction in variable costs (Perrot 1995).

Table 1: 3G Mobile networks in selected economies

	Hong Kong	Japan	Italy	U.K.
Launched in	January, 2004	September 2001	March 2003	March 2003
Service provider(s) and technology	H3G (W-CDMA)	DoCoMo (W-CDMA) KDDI (CDMA)	H3G (W-CDMA)	H3G (W-CDMA)
Number of subscribers (Date)	3,000 (February 1, 2004)	DoCoMo: 2 million (January, 2004) KDDI: 13 million (October 2003)	210,000 (December 2003)	300,000 (January 2004)
ARPU		US\$77.8 (October-December 2001)	Euro 79 (2003)	

Mason’s (1990) study, on the other hand, provided empirical support that product attributes and not cost affect the adoption decision. For rapidly developing technologies such as mobile phones, modularity and upgradability have a strong impact on the diffusion pattern. Rosenberg (1982) found that if a technology is upgradable, it allows a firm to offer product innovations continually that users can integrate into their current product configuration. He argues that users will be reluctant to purchase products that cannot be upgraded and rapid technological change further

increases the reluctance since their investments become obsolete under such conditions (Rosenberg 1982). Upgradability is found to be one of the major factors that influence the selection of a desktop computer. PC buyers are concerned about the possibility of upgrading the processor, adding the memory, availability of drive bays for additional hard drives (Grier and Bryant 1998). Product standardization and modular upgradability also affect and are affected by network externality and compatibility requirements (Kotabe et al. 1996).

Table 2: Supply side factors influencing the diffusion of network goods

Factor	Impact on innovation diffusion	The case of 3G
Expected size of network (NS)	Influence network externality effects	<ul style="list-style-type: none"> • Incompatibility with lower generation phones and PCs. • Too small coverage in some countries
Modularity and Upgradability (MU) Price (PR)	Allows users to integrate innovations continually Influences: <ul style="list-style-type: none"> • Adoption timing • Rate of adoption • Market potential 	Cellular technologies lack consumer-level upgradability <ul style="list-style-type: none"> • Higher prices. Suppliers focused more on reducing variables prices (VPR).
Number of versions/standards (NV)	Multiple versions/standards increase complexity	Uncertain standards and complexities of 3G applications
Degree of supply-abundance (SA) Depth and breadth (DB) Industry-government coordination (IG)	Slow growing and fast declining diffusion pattern under supply restriction. Increase choice and value Inadequate coordination leads to failure.	Shortage of 3G handsets 3G has a wide breadth of uses <ul style="list-style-type: none"> • Market fragmentation • Supply-side crowd • Unrealistic expectations of governments H3G’s confused positioning
Communication/promotion (CP) Other characteristics (OC)	Positioning influences attractiveness.	Size, battery life, reliability of devices and network, etc. H3G’s confused positioning
Communication/promotion (CP) Other characteristics (OC)	Positioning influences attractiveness.	Size, battery life, reliability of devices and network, etc.

Modular upgradability and its benefits are also illustrated by Sun’s SPARCstation 10 series. SPARCstation enables users to increase system performance by plugging in faster microprocessors and memory modules as they become available. SPARCstation 10 series can also be upgraded easily to multiple processing and other future technologies. The SPARCstation 10 provided integrated services digital

network (ISDN) capabilities as a built-in feature. Moreover, SPARC modules and bus architecture can accommodate advances in microprocessors technology. Sun also offered “no penalty upgrades” to its users--the price of an upgrade plus user investment in an existing system would be equal to the price of a new system (Garud and Kumaraswamy 1993).

Inadequate coordination among firms and governments is another source of failure of a new technology (Lin 2003). The failure will also imply the inappropriateness of the industrial policy to other firms and they can avoid that failure by not following the policy (Lin 2003). The diffusion pattern of telecom-related products is also a function of the focus of national industrial and technological policies on fostering and strengthening this sector (Beise 2001, p. 263).

The degree of supply-abundance also influences the shape of diffusion curves. For instance, Simon and Sebastian (1987) found that a diffusion curve exhibits slow growing and fast declining (that is, negatively skewed) pattern under conditions of supply restriction.

Past research has also found that variety and uncertainty of multiple versions increase complexity (Frizelle 1998). A technology's possibility of forking into numerous versions increases the uncertainty. For instance, UNIX in the 1970s was broken into several proprietary formats (Kogut and Metiu 2001), which increased its complexity.

In the case of radical innovations, firms with greater depth and breadth in their product portfolio tend to be more successful (Sorescu et al. 2003). Broader and deeper product portfolio increases consumer choice and hence value of technologies offered by a company. For instance, breadth of product line is found to be positively related to consumers' workstation purchase decisions (Kaplan 1991).

Suppliers can also influence the adoption rate of a product by communicating ideas through advertising and promotion (Narasimhan and Mendez 2001; Rogers 1995) and setting fashions (Abrahamson 1996). Like all other products underpositioning, overpositioning, confused positioning and doubtful positioning (Kotler 2003, p. 311) will slow down the adoption rate.

Table 3: A two-dimensional representation of supply side factors influencing 3G diffusion

	<i>Present Utility</i>	<i>Expected Future Utility</i>
<i>Device Specific</i>	PR, DB, CP, NV, OC (Size, battery life, reliability of device, etc.).	MU, NV, FPR, OC (Size, battery life, reliability of device, etc.).
<i>Network Specific</i>	PR, DB, CP, OC (e.g., reliability of network).	NS, FPR, OC (e.g., reliability of network).

3. SUPPLY-SIDE FACTORS INFLUENCING THE DIFFUSION OF 3G MOBILE COMMUNICATIONS

3G cellular technologies are characterized by different types of network externality compared to lower generations. Whereas second generation (2G) phones could communicate with first generation (1G) phones

as well as with fixed phones, this is not the case with 3G phones. Only 3G devices can exploit combined voice/picture capabilities of 3G networks and lower generation mobile phones and many PCs could be excluded from such application networks (Lehrer et al. 2002). Limited geographical coverage in most countries has exacerbated the problem. Lack of coverage hindered the initial success of FOMA in Japan, NTT DoCoMo's 3G service launched in October 2001. In Italy and the U.K., H3G's coverage started with half the population (Norton, Leslie P 2002). Even by the early 2004, H3G offered 3G services to only 55% of the Italian population (Taaffe 2004).

In a survey conducted by *ChangeWave Alliance*, the Number 1 concern about 3G services of 18% of the respondents was too expensive services and that of another 4% was too expensive handsets. Consider the prices of 3G handsets. In Korea, SK Telecom and KTF had set the price of 3G handsets from US\$581-\$830 in 2003(Business CustomWire, December 28, 2003). In U.K., immediately after the launch of 3G services,H3G had to reduce the price of its handsets by half to pounds 199 which was still much higher compared with the "free" phones offered by 2G operators to those signing up to their monthly plans (Economist.com 2003).

Not only the fixed costs but also variables costs of 3G are expensive. Higher average revenue per user (ARPU) of 3G services reflects unattractive variable costs. In Italy, for instance, ARPU from 3G services was Euro 79 per month in 2003 which compares with 2G ARPU of euro27 in Europe (Taaffe 2004). In Japan, NTT DoCoMo reported that the mean 3G revenue per user in April – June, 2002 was \$65.43 a month (Georgi 2002). In July 2003, 3's UK customers spent an average of £45 a month -much higher than its rivals' £25-£35 (Payton and Bowen 2003).

In an attempt to make 3G services more attractive, the operators have reduced variables costs for some attributes. In U.K. a new tariff introduced by Hutchison 3G in June 2003, offered subscribers 500 minutes for \$40 per month, which compares with the plans of the incumbent 2G operators which averages \$72 for 400 minutes (The Economist, September 20, 2003). In Hong Kong, Hutchison offers three basic service plans to customers at monthly charges of US\$34 (Business CustomWire, January 19, 2004). To stimulate 2G consumers to sign up, NTT DoCoMo cut 3G fees by up to 55% for existing subscribers (Kakenaka 2002).

The wide breadth of uses of 3G has not translated into improvement on each dimension. In the *ChangeWave Alliance* survey, 4% respondents were concerned about current technology's capability to carry high-quality calls (Business CustomWire, November 19, 2003). Subscribers complain of lost calls, bulky handsets that

were awkward to handle and inadequate battery life. In Italy, 15% of early customers took their handsets back (Payton and Bowen 2003).

H3G's pre-launch brand-building campaign, designed to "cultivate curiosity" could not achieve its goal (Payton and Bowen 2003). Hutchison also made too many claims and too frequent change of the positioning of its brand, resulting in confused positioning. Hutchison, for instance, first launched a huge advertising campaign promoting the video capabilities and high-speed internet access of 3G handsets. In September 2003, Hutchison started promoting its 3G services as the cheapest ways to make voice calls instead of touting advanced features such as videoconferencing and Web-browsing (The Economist, September 20, 2003, Ramstad et al. 2003). In fact, the killer application of 3G has become voice communications (Payton and Bowen 2003).

In the ChangeWave Alliance survey, the number 1 concern about 3G services of 17% of the respondents was the reliability of technology and equipment (Business CustomWire, November 19, 2003). Many 3G subscribers are reported to carry their 2G handsets as back-ups. Hutchison 3G mobile in Australia received several complaints of calls dropping out not being able to deliver video and Internet service on the 3G handsets (Business CustomWire, December 28, 2003).

The slow growing diffusion pattern of 3G is can also be attributed to the shortage of 3G ready handsets (Business CustomWire, December 2, 2003). Hutchison also failed to launch the pre-pay service due the lack of suitable handsets. Vendors like Nokia, Motorola, and Ericsson could not ease the supply-side bottlenecks to ensure the availability 3G ready handsets for W-CDMA services. Even 2.5G handsets were seriously delayed by Nokia (Reinhardt et al. 2001).

Many operators around the world enjoy supplier-level upgradability in the cellular network. Economies with CDMA-based 2G networks such as South Korea, Brazil, India and Mexico can easily upgrade to 3G in the form of CDMA2000 (The Economist 2002). Thanks to its CDMA-based 2G standard, South Korea was able to launch the 3G network as early as in 2000. Consumer benefit of the supplier-level upgradability in the cellular network, however, has been minimal. Consumers, for instance, are required to buy a new handset for 3G services. 2G handsets thus cannot be upgraded. There is thus no "no penalty upgrades" (Garud and Kumaraswamy 1993) in cellular networks.

In the 3G mobile industry, there seems to be the lack of proper coordination among governments and mobile players. First, the lack of coordination resulted in market fragmentation. The European Commission for instance, allowed individual governments to make their own decision on 3G spectrum. Such move converted

otherwise homogeneous European cellular market into different licensing regimes, networks and services (Wallage 2000).

Second, cellular operators got into heavy debt because of expensive 3G licenses postponed 3G investments. For instance, Vodafone paid about £ 15 billion to acquire 3G mobile licenses in the U.K., Germany, Italy, Holland and Spain, which doubled its debt level to GBP 13.2 billion in November from GBP 6.64 billion in March 2000 (Naik 2000). Many operators asked governments to ease various requirements and reduce costs (Latour 2002). A petition signed by 2000 European businessmen in April 2001 asked governments in Germany, the U.K. Italy and other countries to return 3G auction proceeds to operators and conduct new auctions (Ryan 2001).

Third, the lack of government-industry coordination also resulted in supply-side crowd in the 3G industry. For instance, consider the European 3G cellular market. In September 2002, Germany had four 3G competitors, the U.K. had six, the Netherlands, Sweden, Finland and Italy each had five, and Spain had four (Latour and Delaney 2002).

These problems are compounded by a lack of policies to foster 3G sector and unrealistic expectations from national governments. For instance, Hutchison faced problem in getting the building permits needed to set up base stations for 3G in Sweden. The Swedish government also specified the requirement of 3G coverage of 99.9% of the Swedish population (Latour 2002).

REFERENCES

1. Abrahamson, E., "Management fashion", *Academy of Management Review*, 21 (1), 254-85, 1996.
2. Beise, M. *Lead markets: Country-specific success factors of the global diffusion of innovations*, Heidelberg: Physica-Verlag, 2001.
3. Economist.com, "The generation game", May 30, p. 1, 2003.
4. Garud, R. and A. Kumaraswamy "Changing Competitive Dynamics In Network Industries: An Exploration Of Sun Microsystems' Open Systems Strategy", *Strategic Management Journal*, 14(5), 351-69, 1993.
5. Gatignon, Hubert and Thomas S Robertson "A Propositional Inventory for New Diffusion Research", *Journal of Consumer Research*, 11, 849-867, 1985.
6. Grier, S.L. and L.W. Bryant, "The case for desktops", *Communications of the ACM*. 41 (1), 70-71, 1998.
7. Jain, Dipak and Ram C. Rao, "Effects of Price on the Demand for Durables: Modeling, Estimation, and Finding," *Journal of Business and Economic Statistics*, 8 (2), 163-170, 1990.
8. Kakenaka, Kiyoshi, "DoCoMo 3G APRU slips 7.5 pct in April-June," *Rueters*, 08/02/02.

9. Kamakura W.A. and S.K. Balasubramanian, "Long-Term View of the Diffusion of Durables," *International Journal of Research in Marketing*, 5, 1-13, 1988.
10. Kaplan, G., "A guide to engineering workstations", *IEEE Spectrum*, 28, Apr, 32-70, 1991
11. Kalish S. and G.L. Lilien, "A Market Entry Timing Model for New Technologies," *Management Science*, 32, February, 94-205, 1996.
12. Katz, M. and C. Shapiro, "Systems Competition and Network Effects", *Journal of Economic Perspectives* 8 (2), 93-115, 1996.
13. Kotabe, Masaaki, Sahay, Arvind, Aulakh, Preet S., "Emerging role of technology licensing in the development of global product strategy: Conceptual framework and research propositions", *Journal of Marketing*, 60 (1), 73-88, 1996.
14. Kotler, Philip. *Marketing Management*, Eleventh edition, Prentice Hall, 2003.
15. Latour, Almar, "Hutchison Unit to Seek Rule Change --- Mobile-Phone Company Wants Sweden to Amend 3G Licensing Regulations," *Wall Street Journal*, Oct 30, p. B.2, 2002.
16. Latour, A. and K. J. Delaney, "Even If MobilCom Bites Dust, Europe Still Has Wireless Glut", *Wall Street Journal*. September 16, p. B.4 2002.
17. Lin, Justin Yifu, "Development strategy, viability, and economic convergence", *Economic Development and Cultural Change*, 51(2), 277-308, 2003.
18. Mason, C.H., New Product Entries and Product Class Demand, *Marketing Science*, March, pp 58-73, 1990.
19. Naik, Gautam, "Vodafone Now Looks Toward Asia --- Strong Earnings Out, British Mobile Giant Wants Presence There", *Wall Street Journal*, Nov 15, pg. A.21, 2000.
20. Narasimhan, Ram and David Mendez, "Strategic aspects of quality: A theoretical analysis," *Production and Operations Management*, 10(4), 514-26, 2001.
21. Perrot, A., "Ouverture a la concurrence dans les reseaux. L'approche strategique de l'economie des reseaux", *Economie et prevision* 119, 59-71, 1995.
22. Ramstad, Evan, Chan Ka Sing and David Pringle, "Hutchison Sticks To Target for 3G; Net Rises by 2%", *Wall Street Journal*, Aug 22, pg. B.4
23. Reinhardt, Andy, Stephen Baker and William Echikson, "All That Money on 3g--and for What?: A simpler, cheaper wireless Net standard could win out", *Business Week*, March 26, p. 60.
24. Rosenberg, Nathan, *Inside the Black Box: Technology and Economics*, Cambridge (UK): Cambridge University Press, 1982.
25. Ryan, Vincent, "Reversal of spectrum fortunes", *Telephony*, April 2, pp. 14-15, 2001.
26. Simon, H. and K. Sebastian, "Diffusion and Advertising: The German Telephone Campaign," *Management Science*, 33 (April), 451-466, 1987.
27. Taaffe, Ouida, "H3G Italy: 3G with styles?", *Telecommunications International*. 38 (1), p. 22, 2004.
28. The Economist "Business: 3G by any other name, Mobile telecoms", Jan 12, p. 72, 2002.
29. The Economist, "The shunning of 3G, September 20, 2003, p. 59, 2003.
30. Wallage, Steve, "European 3G auctions: There must be a better way", *Electronic Business*, 26 (7), p. 30, 2000.

The Implementation of Signing e-Document by Using the Wireless Identity Module in Cellular Phone

Chengyuan Ku*, Yenfang Ho, Yiwen Chang

Department of Information Management, National Chung Cheng University,
 160, San-Hsing, Min-Hsiung, Chia-Yi, Taiwan, China
 Phone : 886-5-2720411 ext. 34603, Fax : 886-5-2721501
 {cooperku, bobo, azure}@mis.ccu.edu.tw

ABSTRACT

Traditionally, the document was recorded on the papers and signed by related personnel to differentiate the responsibility. As the technology of communication is evolved quickly, it makes people connect each other by many different ways. Now, the e-document can be transferred on the network and e-signed with proper tools. However, it still needs people to stay in front of desktop to send and receive e-document. In this paper, we propose a way to send and receive the urgent e-document with cellular phone. For the sake of safety and non-repudiation, the users need to e-sign the document after receiving and before sending our e-document using the embedded wireless identity module (WIM) inside the SIM cards. The use of WIM in the SIM card to sign e-document can make the signing process more secure and reduce the possibility of argument. A prototype of this system was implemented. As can be imagined, the system will provide convenience and security for top managers to handle urgent events and e-document on the move.

Keywords: Smart card, wireless identify module (WIM), electronic document (e-document), mobile office

1. INTRODUCTION

The computers and networks give lots of convenience to us. We can process our e-document very quickly no matter where we are as long as we have the access of Internet. Now, it is convenient to use desktops to get the access of Internet. However, many top managers need to travel very often to handle the remote business. If the notebook is not suitable to be carried with in the long trip and the manager needs to handle the urgent document on the move, the cellular phone or personal digital assistant (PDA) could be a convenient tool to take the initial step. For the consideration of safety and non-repudiation, the proposed system was designed to allow the mobile manager to handle e-document using digital signature through the mechanism of WIM inside the SIM card. The smart card is used to prevent from intrusion by hackers and to show the users' identification to e-document servers. The benefits of using a smart card include [1] :

- | increased security,
- | potential user mobility, and
- | sequential access to one machine by multiple users.

In 1997, an industry organization named the WAP Forum was founded to develop technical standards to bridge the gap between the mobile and Internet networks. The development of WAP brought the wireless world closer to the Internet through a set of specifications utilizing technologies that enhance the experience of wireless users [14]. The first draft of

specifications was published in February 1998. The subsequent release of WAP 2.0 with additional features allows WAP to support improvements in wireless devices and Internet content technologies [14]. WAP can be used to define both communication protocols and application environments. The protocol optimizes the standard Web protocols for use in the low bandwidth, high latency conditions prevalent in wireless networks. Various enhancements to the session, transaction, security, and transport layers make HTTP (Hypertext Transfer Protocol) functionality better suited to the wireless network environment [20]. WAP security functionality includes the Wireless Transport Layer Security (WTLS) and application level security [12]. The WTLS has three classifications of security and provides different functionality as presented in Figure 1.

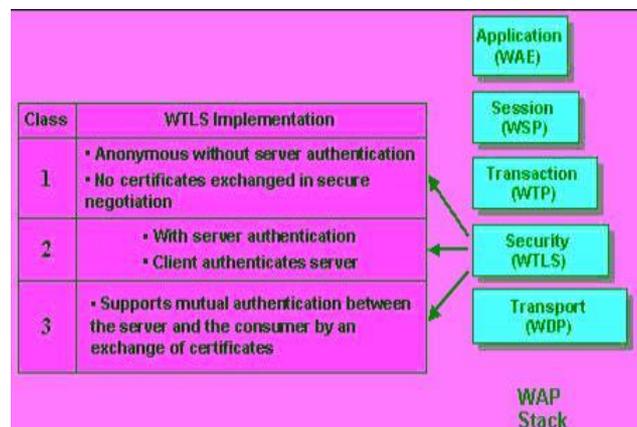


Figure 1. The Security Classification of WTLS [13]

* Corresponding Author

For optimal security, some parts of related functionality need to be performed by a tamper-resistant device [12], especially for level 3 of WTLS. The Wireless Identity Module (WIM) is one of the WAP security function. The WIM is used to perform Wireless Transport Layer Security (WTLS) and application level security functions, and especially to store and process information needed for user identification and authentication [12]. The smart card is the best tamper-resistant equipment. So, our design is to use a cellular phone to provide the better mobility, and the digital signature and mutual authentication will be done by using the WIM embedded in the SIM card. This is the so-called SWIM card. Using the SWIM card to compute the digital signature and authenticate client and server, it can provide the security, mobility and also non-repudiation for e-document transactions.

2. RELATED TECHNOLOGY

2.1 SMART CARD

There are two kinds of smart cards. One is storage card and the other is processor card. People are already accustomed to some applications where a smart card is used just to store data, such as telephone cards and health insurance cards. Currently, the most popular application for processor cards is the SIM card within cellular phone [11]. There are many kinds of applications, which are related to the smart cards. Table 1 presents these applications.

Table 1. The Market Share of the Smart Card [16]

Domain Market	Market Share
Financial Services	16%
Government	13%
Communications	12%
Healthcare	11%
Retail	9%
Transport	8%
Services	7%
Education	7%
Insurance	6%
Utilities	6%
Manufacturing	4%

Because the smart card is a tamper-resistant device, it is widely used for authentication, making payments, or issuing digital signatures. The smart card can protect from physical attack and provide mechanisms against light or abnormal voltage detection, so the security of embedded data is ensured through the use of various coding techniques [11].

To exchange data between the smart card and the terminal, APDUs (Application Protocol Data Units) should be used [9]. The APDU denotes an internationally standardized data unit in the application layer [9]. Table 2 is the format of the command APDU and table 3 is the format of the response APDU.

Table 2. Structure of Command APDU [2, 9, 12]

Mandatory Header				Optional body		
CLA	INS	P1	P2	Lc	Data field	Le

Table 3. Structure of Response APDU [2, 9, 12]

Optional body	Mandatory Trailer	
Data field	SW1	SW2

A command APDU, which is sent by the card reader to the smart card, is composed of a header and a body. The header consists of the four elements [2][9]:

- u CLA (class of instruction): the class byte is used to identify applications and their specific command sets.
- u INS (instruction code): which encodes the actual command.
- u P1, P2 (parameters 1, 2): these two bytes are primarily used to provide more information about the command selected by the instruction byte.

The section following the header is the body, which can be dispensed with except for a length specification. The elements of the body are described as follow [2][9]:

- u Lc: Lc is the length of the data section sent to the card.
- u Data field: contains the data associated with the command that are sent to the card.
- u Le: it specifies the length of the data section to be sent back from the card.

The response APDU, which is sent by the smart card in reply to a command APDU, consists of an optional body and a mandatory trailer, as shown in Table 3. The length of the data field may be zero, regardless of the value specified in the command APDU [9]. SW1 and SW2 are the status words to encode the response to the command [2].

2.2 ELECTRONIC OFFICIAL DOCUMENT

There are many disadvantages of manipulations of traditional official document. It may take much of time to deliver the official document and it may be also difficult to reserve the physical document or retrieve it. So our government planed to computerize the official document system and deliver the official document on the Internet since 1993. The electronic document use XML (eXtensible Markup Language) as the national standards. Through the usage of the XML, electronic official document can be exchanged automatically and it is also very easy to be retrieved and maintained. The following Figure 2 describes the infrastructure of e-document exchange. First, the sender should build the official document via XML standard and then use the

in security related commands. Each SE specifies references to the cryptographic algorithms to be executed, the modes of operation, the keys to be used and any additional data needed by a security mechanism [12]. After storing the SE, we should use MSE-SET command to set the location of key of the current SE. Figure 6 illustrates the procedure of computing digital signature.

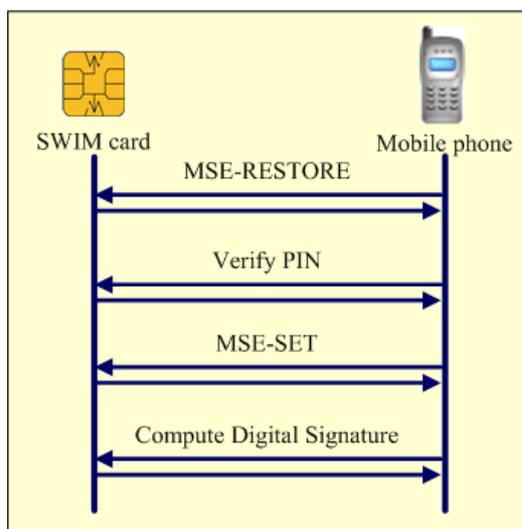


Figure 6. Procedure of Computing Digital Signature [14]

Please set the paper size as A4 (29.7cm*21cm). Leave 2.5cm margins at both the top and the bottom of the page, 2cm on both right and left sides. Please write your paper using MS-Word. The Word of Microsoft's office 97 and 2000 is strongly preferred. If you write the paper by using other versions' Word, please give clear indication of what version of Word you use when you submit the paper by email.

4. PROTOTYPE

The system has two parts, which are the web server and the mobile client. We build the electronic documents with web interface on the server and communicate with SWIM card on mobile client. Therefore we can e-sign the remote digital documents on the web server through the WIM modules in the SWIM card.

Microsoft windows 2000 was chose as the development platform and the JSP (Java Server Pages) and Tomcat 5.0 had been used to implement the website. The development of the mobile client is based on J2ME. First, the user has to login the website of the official document. After the user login, he/she can fill out the electronic form of official document as shown in Figure 7.



Figure 7. Electronic Official Document



Figure 8. Compute Digital Signature

After filling out the form, the user will select the button to compute the digital signature as shown in Figure 8. First, the SHA-1 (Secure Hash Algorithm) is used to compute the digest of official document. Then the web server will transfer this hash value to the mobile client and the mobile phone will invoke the related service primitive to compute the digital signature. If needed, the user can also verify the digital signature through the SWIM card and the card will return the verify status as shown in Figures 9 and 10.



Figure 9. Verify Signature



Figure 10. The Screen of Verify Status

When the user receives the electronic official document, he/she can read and e-sign the receipt as shown in Figures 11 and 12. Therefore the sender of official document knows that his/her document arrives intact.



Figure 11. The letter of receiver



Figure 12. Signing the document while receiving

5. SECURITY ANALYSIS

The security of this system is based on the secure mechanism of the SWIM. We will examine the system security based on the information security

standards- ISO/IEC 10181 [4], which are defined by International Organization for Standardization (ISO) and International Electrotechnical Commission. The ISO/IEC 10181 information security standard can be divided into four part, they are Confidentiality, Authentication, Integrity and Non-repudiation. We will ignore the confidentiality because we assume the document is not confidential.

u Authentication

Authentication means that the individual is who he or she claims to be. Our system uses SWIM card to authenticate the user. The user should enter his/her PIN code in order to authenticate himself/herself. If the card owner entered the wrong PIN code more than three times, the SWIM card will be blocked and the original PIN code will also be useless.

u Integrity

The system should prevent the modification of document when it is transmitted. Our system keeps the digest in the web server. If these digests are not consistent then server will ask the mobile client to send it again.

u Non-repudiation

Non-repudiation means that the sender of official document can't deny ever transmitting this document and the receiver can't deny ever getting it. We use the digital signature to prevent the sender or receiver from repudiating their transaction.

6. VARIABLES AND EQUATIONS

In this paper, we propose the system of electronic official document, which can be processed via the cellular phone. In the future, the 3G communication network will provide higher data rate, then the capability of processing multimedia document can be embedded in our system. By adopting the mobility of cellular phones and high data transfer rate, we will go one-step further to the ideal mobile office.

REFERENCES

- [1] Chadwick, D.; "Smart cards aren't always the smart choice", *Computer*, Vol. 32, pp. 142-143, 1999.
- [2] Chen, Z.; "Java Card Technology for Smart Cards: architecture and programmer's guide", Addison Wesley, 1st edition, 2002.
- [3] F. P. a. M. R.; Kenneth G. Paterson; "Smart Card and the Associated Infrastructure Problem", *Information Security Technical Report*, Vol.17, pp. 20-29, 2002.
- [4] ISO/IEC International Standard, "Information Technology- Open Systems Interconnection- Security Frameworks for Open Systems: 10181", Part1-Part7, 1996.
- [5] Mobile Electronic Transaction, "PTD Definition Version 2.0", Oct. 2002.
- [6] Nelson, Matthew; "56-bit DES algorithm broken in record time", *Infoworld*, January 25, 1999, p. 8.

- [7] Nokia Corporation, "How to utilize Wallet and WIM in mobile transaction services", 2002.
- [8] Peyravian, Mohammad; Matyas, Stephen M.; Roginsky, Allen; Zunic, Nevenko; "Generation of RSA Keys That Are Guaranteed to be Unique for Each User", *Computers and Security*, Vol.19, Issue: 3, March 1, 2000, pp. 282-288
- [9] Rankl, W.; Effing, W.; "Smart Card Handbook", Wiley, 2nd edition, 2001
- [10] Stallings, William; "Cryptography and Network Security Principles and Practices", Pearson Education, 3rd edition, 2003.
- [11] Scheuermann, D.; "The smartcard as a mobile security device", *Electronics & Communication Engineering Journal*, Vol.14, Issue: 5, Oct. 2002, Page(s): 205 –210.
- [12] WAP Forum, "Wireless Identity Module", Version 12-July-2001.
- [13] Douglas A. Kuhlman; "Practical RSA vs. ECC Speed Comparisons", 2002.[14] WAP Forum <http://www.wapforum.org>
- [15] OpenCard <http://www.opencard.org>
- [16] FIND <http://www.find.org.tw>
- [17] International Organization for Standardization <http://www.iso.ch>
- [18] International Electrotechnical Commission <http://www.iec.ch>
- [19] SchlumbergerSema <http://www.axalto.com>
- [20] Shin-Yuan Hung, Cheng-Yuan Ku, and Chia-Ming Chang, "Critical Factors of the Adoption of WAP Services: An Empirical Study," *Electronic Commerce Research and Applications*, vol. 2, no. 1, pp. 42-60, 2003.

The Quest for Mobility: Designing Enterprise Application Framework for M-Business Practices

Chianhsueng Chao

Department of Information Management, National University of Kaohsiung,
Kaohsiung, Taiwan, China
cchao@nuk.edu.tw

ABSTRACT

With the advent of the “road warrior,” a growing part of the workforce is mobile and using all sorts of mobile devices to stay in touch and transact business. As the global economy shifts toward the mobile economy, enterprises need to be progressively more flexible and globalize. Mobile businesses open up new opportunities for innovative enterprises and give them new means of communications with customers and employees. In a changing business landscape, mobile business addresses new customer channels and integration challenges. The current transformation is simply the movement of e-business to a mobile environment. It is still a developing concept as are the business models that support it. The usefulness of the mobile channels will be largely driven by new enterprise applications that enhance the overall customer values. This paper provides a broad discussion on the movement of mobile integration strategies. Several issues will be addressed, such as value chain, the data access of mobile computing, m-business application framework, and the future development of mobile computing. All of these efforts attempt to provide an overview and schematics for the integration of modern e-business application strategies into future m-business practices. This paper will show that any proposed system or strategy must recognize the primary value and mechanism of how people “work,” and technological solutions must be devised in order to facilitate people who conduct business. With mobile technologies, enterprise applications will go beyond the four walls of organizations to a workforce on the move.

1. THE PARADIGM SHIFT

The challenge of global marketplace and new emerging technologies are the driving forces for enterprise to optimize their business strategies and operations frequently. Customer preferences and needs change constantly, resulting in pressingly intense competitions. Time sensitive real-time transaction is becoming the common denominator for successful businesses. Information also needed to be available to anyone from anywhere at anytime. There are challenges along the path of business evolutions, however, the innovated enterprises seize every opportunity to reinvent themselves in order to fit into the global spectrum of business. The Internet is bringing profound impact to the business world and has enabled a new way of conducting commerce. To compete in the emerging digital economy, enterprise need to change their business models, rethink the way they work and forge new relationships with their trading partners and customers.

Today, mobile businesses open up new opportunities for innovated enterprises and give them new means of communications with inside/outside stakeholders. In a changing business landscape, mobile business addresses new customer channels and integration challenges. The e-business comes a long way from hype to mature. The mobile solutions will build on the top of the existing e-business investments. The m-business is facilitated by the integration of the Internet, e-business, and the wireless world where customers can go online with any device at anytime from anywhere. Enterprises focus on

how to use mobile technology to provide customers with rich service experiences and greater satisfactions. This is a more real picture about business, because business opportunities occur in anywhere at anytime. To embrace this big vision, a new value chain is emerging which focuses on the mobile devices, access methodology, and enterprise contents. Mobile applications are beginning to play a central role in enabling real-time supply chain. Enterprise must design a mobile application framework for both customers and their suppliers. The usefulness of the mobile channels will be largely driven by new enterprise applications that enhance the overall customer values. Seeing m-business processes in terms of a net value added chain is a good approach for visualizing current and future competitive advantages. The creation of these values depends on the enterprise’s ability to coordinate and link those internal/external activities efficiently.

2. THE WIRELESS LANDSCAPE

The m-business is facilitated by the integration of the Internet, e-business, and the wireless world where people can go online with any device at anytime from anywhere [1]. All mobile workforce demands mobile computing and access to e-commerce resources regardless of the medium used to gain access. However, to create multi-channel and multi-technology solutions in a fast-moving and technologically indefinite mobile environment is difficult and extremely risky. A number of mobile devices, closed networks, disparate carrier/server systems, protocols, and solutions characterize today’s

wireless landscape. Currently, accessing the web through a wireless device is different than accessing the web through wired connections. The fastest way to access the Internet on the move is through the IEEE 802.11b (Wi-Fi) standard. Overall, 802.11 (Wi-Fi) in its various forms (a, b, e and g) has emerged as the winner in wireless LAN technology [2]. However, the problem with wireless LAN-based access is that the access points or base stations have a very short range, which means they're restricted to a few "hotspots," such as hotels, conference centers, and airport. The situation gets more complex when the wireless LAN is mobile.

2.1 The Wireless-to-Wired Network Infrastructure

The development of the cellular network not only enabled the advancement of mobile communications but

also provided the playgrounds for mobile commerce. The term cellular comes from the honeycomb shape of an area into which a covered region is divided. Cells are base stations transmitting over small geographic areas. A frequency spectrum is divided into blocks of frequencies and those blocks are assigned to the cells. As a mobile unit passes from one cell site to the next, the call is handed off from one transceiver to another. When the service area grows with more users, the approach is to split an area into smaller ones. In this way, the urban centers can be split into as many areas as necessary to provide service in heavy traffic regions. Therefore, less expensive cells are then deployed to cover the remote rural areas. The base station (BS) host several telecommunication equipment at the transmit/receive antenna's base, the center of a cell.

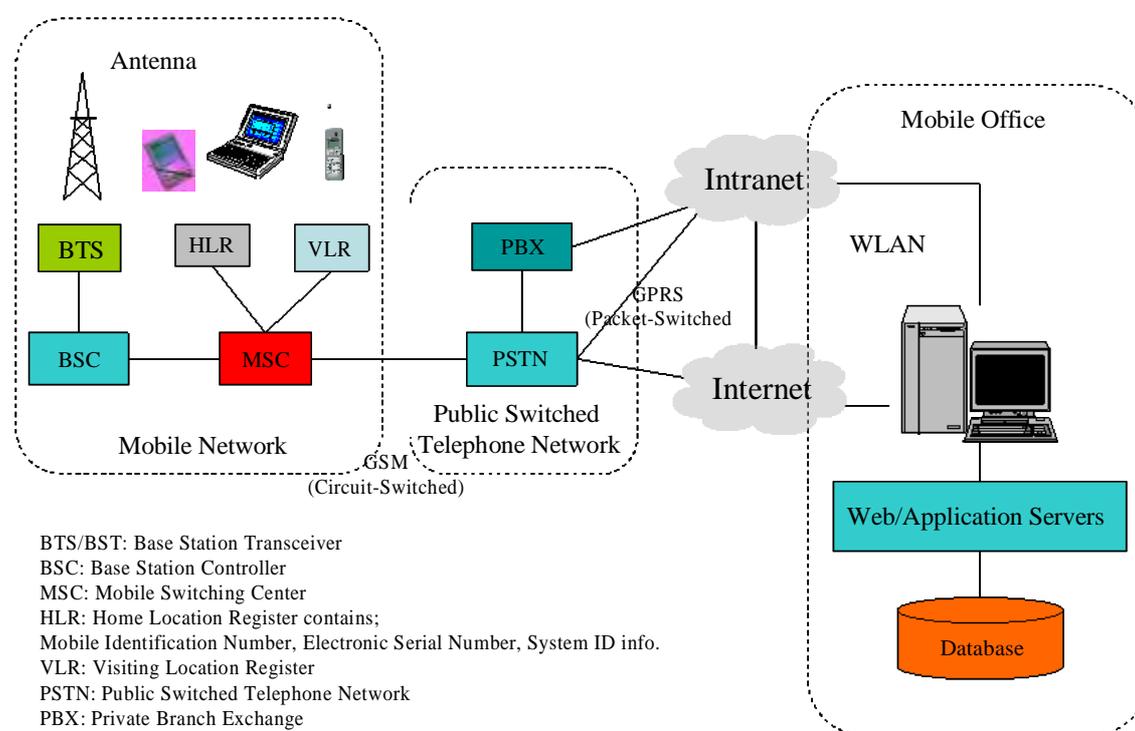


Figure 1 The Wireless-Wired Landscape

The base station in Figure 1 consists of two major electronic units known as the base station transceiver (BST) and the base station controller (BSC). The BSTs are full-duplex transmitter and receiver units (transceivers) that signal to the mobile units and receive signals from them. The BST serves one cell in the cellular network and contains one or more transceivers. The BSC controls the transmission and reception of signals of the BST and switches calls between the BST and mobile units. The BS is also responsible for administration, call processing, maintenance, handoff, and audio signal compression and decompression as directed by the mobile switching center (MSC). Each BS is controlled by the MSC. The BS is connected to the MSC using voice trunks and data links. The MSC serves

as a central coordinator and controller for the cellular system and is also an interface between the wireless and wired infrastructure. The MSC is connected to the Public Switched Telephone Network (PSTN) using voice trunks and data links. The MSC can be considered the same as a landline-switching center in the central office (CO) with the enhanced capability to track each mobile unit. A system identification number (SID) is used to identify whether a person is roaming or in the home (home SID) area. If a person makes a call within the home area, then the MSC serving the call is considered the home MSC. If the person is roaming outside the home area, then the MSC serving the call is known as the serving MSC. In that case, the SID of the MSC is known as serving SID.

In addition, there are Home Location Register (HLR) and Visiting Location Register (VLR) databases owned and maintained by the service providers that keep track of subscriber's information. In this way, the mobile users can get on the Internet using technologies, such as Wideband Code Division Multiple Access (WCDMA), which is usually referred to as 3G.

The fourth-generation (4G) cellular services intend to provide mobile data rates of 100Mbps or more. This 4G is an entirely packet-switched network that all network elements are digital with higher bandwidths to provide multimedia services at lower cost. In 4G mobile IP, each cell phone is assigned a permanent "home" IP address, along with a "care-of" address that represents its actual location. When a computer somewhere on the Internet wants to communicate with the cell phone, it first sends a packet to the phone's home address. A directory server on the home network forwards this to the care-of address via a tunnel, as in regular mobile IP. The directory server also sends a message to the computer informing it of the correct care-of address, so future packets can be sent directly. This should enable TCP sessions and HTTP downloads to keep going as users move between different types of networks. The ultimate 4G would be full on-demand capability, Wi-Fi to WAN roaming, and have it as a true IP network. Currently, the Japanese government required all of Japan's ISPs to support IPv6 by 2006, to coincide with the launch of 4G. Countries in Europe and Asia are considering similar regulatory action. NTT DoCoMo (provides i-mode service) also hopes that true mobile broadband will enable them to replace fixed access entirely, and they plan to have a 4G system operational by 2006.

3. DEFINING MOBILE MARKET VALUE

Whenever a technological innovation is introduced, it brings both new business opportunities and new challenges to enterprises. As the global economy shifts toward the mobile economy, enterprises need to be progressively more flexible and globalize. The mobile business strategies will focus on what "value" might be created for the enterprises in the marketplace.

To gain competitiveness and create values, the enterprise should target on redesigning their core business processes in response to key competitive factors. These competitive factors can be derived from the Porter's value chain [3]. The value chain divides the organization into a set of generic functional areas, which can be further divided into a series of value activities. In the value chain, there are two distinct types of functional area: primary and support (Figure 2). Primary activities are concerned with the direct flow of production (such as inbound logistics, operations, outbound logistics, marketing, sales, and service), whereas support activities (firm infrastructure, human resource management, technology, and procurement) support the primary activities and each other. Starting with its generic value chain categories, a firm can subdivide into discrete activities, categorizing those activities that contribute best to its competitive advantage. Porter's version of supply chain management is called a value chain because it focuses on value. The value is measured by the amount customers are willing to pay for an organization's product or service. Primary and support activities are called value activities, and an enterprise will be profitable as long as it creates more value than the cost of performing its value activities [4]. In this way, a value chain is defined and a better organizational structure and business process can be created around those value activities that can most improve an organization's competitive advantage [5].

Porter also recognized linkages outside the enterprise, as they relate to the customer's perception of value. This provides the possibility that one value chain could be linked to another value chain, because one business partner could be the other's customer. This interconnected value chain system can act as a supply chain that encompasses the modern business world, and participating organizations can readily extend their technologies to their partners. The "extended enterprise" aspect enables supply chain integration, more effective outsourcing, and self-service solutions for both internal and external stakeholders [5]. The usefulness of the mobile channels will be largely driven by new enterprise applications that enhance the overall customer values. A business that creates m-business applications can benefit from the first-to-market advantage. As the wireless Internet continues to expand,

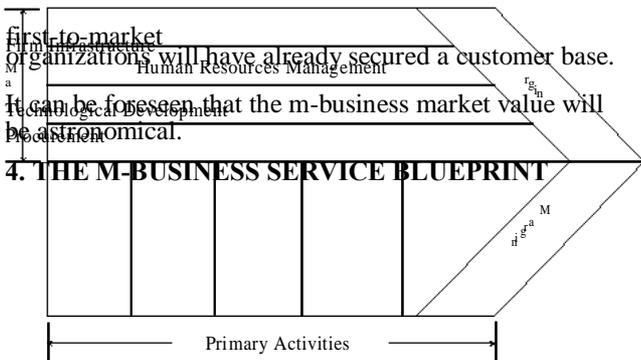


Figure 2 Porter's Value Chain

As mentioned before, seeing m-business processes in terms of a net value added chain is also an approach for visualizing current and identifying future competitive advantages. In short, the creation of these values depends on the enterprise's ability to link and coordinate these internal/external activities efficiently [3]. This

interconnected value chain system can act like a supply chain that encompass the modern business world, and participating organizations can readily extend their technologies to their partners. The m-business supply chain practices are just like e-business supply chain practices with the exception of mobility. This is a more real picture about business, because business opportunities occur in anywhere at anytime.

With the convergence of B2B, B2C, SCM, CRM, ERP, Business Intelligence, and real-time relationship management, the personalization of customer information is now possible. The IP convergence strategy will provide a solution that joins the two worlds of voice communications (telephony) and data traffic together. Technologies, such as interactive voice response (IVR), computer telephony integration (CTI), are beneficial to selling chain management. The fully interactive nature of the Internet changes everything in the world of self-service. The blending of Sales Force Automation (SFA) with customer self-service brings new business opportunities. Relevant information and intelligence support processes all can be aligned and implemented in real-time at anywhere. The application and technological framework of m-business to facilitate

these highly interactive communications will be illustrated in the following section. This framework extends existing business applications such as ERP, CRM, and mobile workforce applications to wireless devices.

5. THE M-BUSINESS SYSTEM ARCHITECTURE

In today's m-business, competing for mobility requires either aligning one's strategy to what the enterprise agility or developing mobile computing capabilities to support a desired supply chain strategy. Mobile business systems must to address the strategic visions of enterprise. A successful mobile business application framework for enterprises should link supply chain management, relationship management, and knowledge management to function in an adaptive way and continue to thrive in the m-business era. The design of a m-business system needs to focus on collaboration. In the m-business system architecture, several dedicated servers and databases are required, because enterprise must capture and retain in a central/distributed data repository the data and information that employees need (Figure 3).

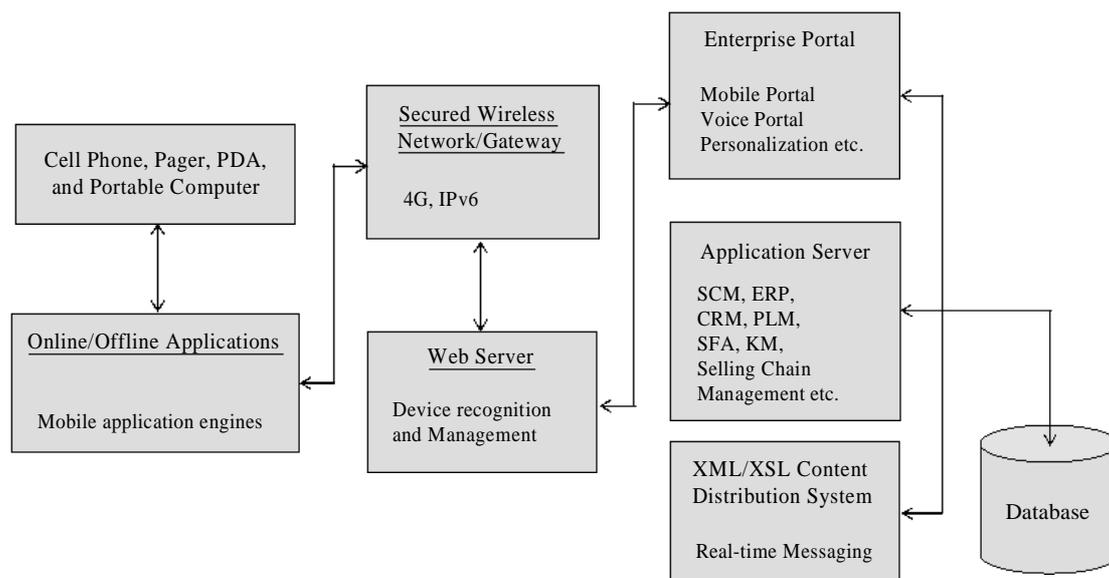


Figure 3 M-Business Application Framework

A selling chain management suite for customer care applications enables new kinds of interactions, new kinds of information, and comprehensive access to all customers and enterprise portals. This customer care can foster joint problem-solving, and convenient self-service. Given the information customers need in the purchasing process. The reasons to stress on 4G and include in this framework are (a) it has wider bandwidths, higher bit rates, (b) it support interactive multimedia services, teleconferencing, wireless Internet, etc., and (c) the global mobility and service portability. For knowledge

management issues, the architecture of IT in the knowledge management system is concerned with organizing and analyzing information in an enterprise's database so this knowledge can be readily available throughout an enterprise. A central/distributed data repository that provides/captures the data and information for employees and executive decision-making is also very important in a knowledge management system. The database system is the driver that consolidates and directs the overall resources of the supply chain to the most mission-critical business

activities. As a result, an m-business supply chain solution must provide supports for the capture and communication of customer demand, as well as enable this demand to automatically trigger business events and initiate process workflow. Transcoding contents and developing enterprise applications to work across several mobile devices and browsers seem to be a complex problem. The key is to focus on the lowest common denominator: the XML (eXtensible Markup Language) and XSL(eXtensible Style sheet Language). The XML supports for developing mobile applications running on multiple devices and platforms [6]. Databases are also working with XML-structured data and tabular relational data. The ability to translate standard HTML/XML content into the various flavors of wireless content is based on the way that the XML/XSL standards operate. Because of these standards, XML is able to function as a meta language, flexible enough to recreate other markup languages and use them as subsets. Technologies such as Java Server Pages (JSP) and Active Server Pages (ASP) are also being extended to support multimode clients. Figure 4 and Figure 5 illustrate some examples of transcoding systems and applications. Enterprises that provide HTML to wireless transcoding solutions translate standard content into XML that reflects all of the information flow and logic of the transcoded content, and by browser detection determines for each mobile user which subset of XML to translate content into [7]. Thus we see that a robust and active transcoding system goes a long way toward reducing complexity and maintaining consistency throughout the enterprise across multiple clients. Effective m-business supply chain solution will need to deliver an accurate and common view of customer demand data as well as any subsequent events, plans, or other business data. This new supply chain framework will offer virtually unlimited business opportunities in the alignment of technologies and processes. A well-designed and well-integrated supply chain for m-business will improve upon existing cost-responsive processes, and have organizational agility in the event of change.

6. CONCLUSION

The emerging mobile technologies with Internet capabilities that connect numerous devices and information portals represent a new phase in enabling intelligent mobile workforces. Multi-channel and multi-technology strategies represent the next frontier. Mobile business addresses new customer channels and integration challenges. Being the central platform for mobile applications, the next generation Internet must provide an infrastructure that allows data transfer to be fast and more cost-efficient. The mobile Internet will open an important new channel for commerce. With mobile technologies, enterprise applications will go beyond the four walls of organizations to a workforce on the move. Employees will be able to access the company’s mission-critical enterprise applications, such

as enterprise resource planning (ERP), customer relationship management (CRM), and supply chain management (SCM). This open enterprise is also extended to suppliers and trading partners so that when customers get in contact with the resources of an enterprise, they also touch upon the resources of the value chain. The usefulness of the mobile channel will be largely driven by these enterprise applications that enhance the overall customer values. With self-service applications in hand, customers will be able to touch the resource of an enterprise. It is the customers who drive the entire value chain (or supply chain), determining what is to be produced, when, and at what price. They will be able to interact with the entire business ecosystem and not just individual enterprise.

Content in XML can use XSL style sheets to dynamically render the content into:

- HTML for Desktop and PDA
- HDML or WML for mobile devices
- Specific XML for B-to-B

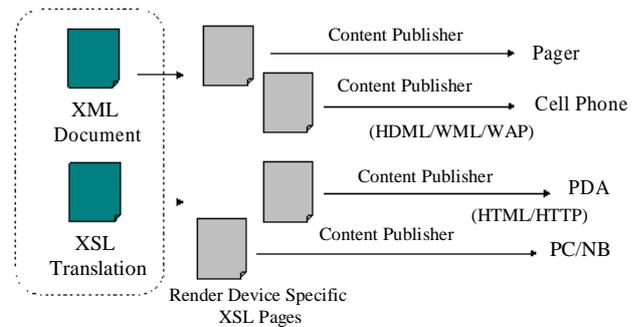


Figure 4 The Wireless-Wired Landscape

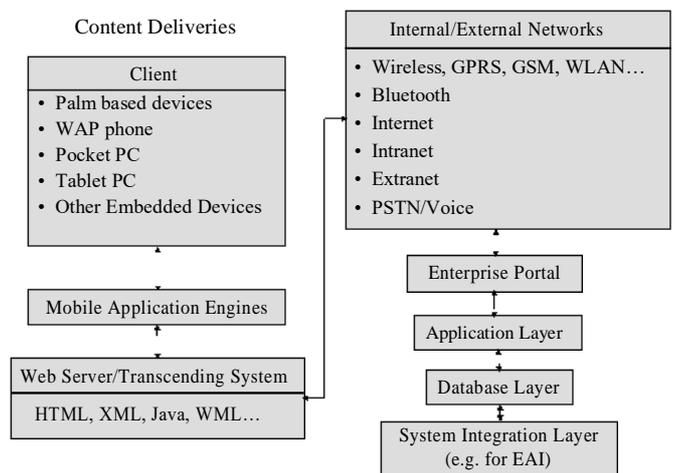


Figure 5 Content Delivery Schemes

REFERENCE

[1] Kalakota, R. and Robinson, M. “M-Business, The Race to Mobility.” McGraw-Hill, 2002.
 [2] Editorial Group of Wireless Advisor Zone “SSL Leads the Way in Wireless Security Technologies.” An

Internet Article at Wireless Advisor Zone, <http://wirelessadvisor.net/>, May 2003.

[3] Porter, M. "Competitive Advantage: Creating and Sustaining Superior Performance." The Free Press, 1985.

[4] Kuglin, F. A. "Customer-Centered Supply Chain Management, A Link-by-Link Guide," AMACOM, a Division of American Management Association. 1998.

[5] Curran, T. A. Ladd, A., and Keller, G. "SAP R/3 Business Blueprint, Understanding Enterprise Supply Chain Management," Prentice Hall, Inc. 2000.

[6] Jerke, N. E. "Building Mobile Applications with Microsoft Tools." An Internet Article at Mobile Development Advisor Zone, <http://wirelessadvisor.net/>, August 2003.

[7] Bergeron, Bryan "The Wireless Web, How to Develop and Execute a Winning Wireless Strategy" McGraw Hill 2001.

The Researches on the Mobile E-Business Implements and Mode in Small and Medium-sized Enterprise

Yingchun Wang^{1,2}, Changtong Wei¹, Zhuang Guan¹, Chunling Sun^{1,2}, Qiong Zhu¹

¹ School of Economics and Management, Tianjin University of Technology, Tianjin 300191, China

² School Management, Tianjin University, Tianjin 300072, China

tjycwang@vip.sina.com

ABSTRACT

The "mobile e-business" is a business activity derived out by the concept of the "e-business", which can finish the specific commercial service function through mobile devices, such as cell-phone, PDA, etc, with the marketing purpose of achieving products and service. Mobile e-business is gradually known by kinds of enterprises, and successively put into practice among them. This paper sets out from analyzing the prospect of mobile e-business and current situation of small and medium-sized enterprise of our country, proposes the technology, implementation means, mode of the mobile e-business for the small and medium-sized enterprise, and compared with the application mode of the large-scale company with its one's own characteristic.

Keywords: the mobile e-business, the small and medium-sized enterprise, implements, mode

1. THE PROSPECT OF THE MOBILE E-BUSINESS

As far back as 2000, the mobile e-business once becomes the focus that everybody looks forward to it. A global investigation indicates, saying 61% of those who are investigated all have their heart filled with hopes: in the near future, they can regard the wireless device as an ordinary payment tool. Analysts who engaged in this investigation predict the mobile commercial market can reach 100 billion dollars in the whole world in 2003. Though the present situation is not optimistic, the rapid development of the mobile e-business need not be queried^[1].

According to the news of "net of PC World", a market survey report of Frost&Sullivan's company shows that, Because of the emerge of a large amount of portable equipment such as cellular telephone, PDA, etc, the mobile e-business after more than two years quiet, will develop at a full speed over the next several years. It is predicted, the market of mobile e-business will reach 25 billion dollars by 2006, which share will account for about 15% of the global online trade market.

Nearly there are 250 million cellular telephone users at present in China, and added about 4 million pre month; the market penetrance is 15%, which is relatively at low levels. It proves that a sizable potentials market still remain to be explored. It is estimated that by 2010, the cellular telephone will exceeds 540 million users, the popularity will reach 40%; In 2002, the value-added telecom service "short message" created over 30 billion. During the Spring Festival of 2003, the user of China Unicom and China Mobile delivers 7 billion pieces of SMS news altogether^[2].

Therefore, to develop the mobile e-business, expand the products and service in the mobile field that the

enterprise utilized, raising the income based on mobile service, is a piece of forward position for the enterprise's development, which will bring the enormous profit space for the enterprise.

2. THE NECESSITY THAT THE SMALL AND MEDIUM-SIZED ENTERPRISE DEVELOPS THE E-BUSINESS

Headings are numbered and capitalized. All major headings are centered in bold in 10 fonts. Do not put a period after the text of the heading. Leave one line above a major heading, and one line clear below before the start of the next paragraph or second-level heading.

As the elementary cell of national economy, the quantity small and medium-sized enterprise in our country exceeds 99% of all the enterprises, Small and medium-sized industrial enterprise account the 60% of total gross output value, the sales income accounts for 57% of total value. So, to some extent, the level of the informationization of small and medium-sized enterprise has reflected the current situation of the informationization of enterprise in China. Their demand for informationization represents an important respect of information-based market demand of Chinese enterprise too^[3].

Seen from resource of the hardware, CCID investigation has shown, in the enterprise of our country, the having rates of the desktop, printer are nearly 100%, and the having rate of the server exceeds 50% too. According to the statistics of New network ChinaDNS, now, the enterprise which participates in the e-business is only 22.3%, and among them, the incidence of net inquiring reaches 72.9%, issuing the information on the net reaches 71.4%, which are relatively high, but it isn't universally applied in supplying chain integrating, net paying, distribution channel, etc. Among them, Web

sites which have online-paying system are less than 1/3, and web sites which carry on e-business are 1/7.

According to a large amount of data above, the Informationization level of the small and medium-sized enterprise of our country remains to be strengthened, The level that the enterprise uses the e-business still needs to be improved further. The enterprise utilizes the advanced means of mobile network, extend the concept of the e-business to all kinds of terminal stations, which will make enterprise as closer to customers as possible, make the customers know the enterprise and its products, make the enterprise be in the front row throughout in the fierce market competition, be keeping vigorous.

3. THE APPLICATION KINDS OF THE MOBILE BUSINESS

At present, the enterprise implementing the mobile e-business is to carry on the commercial activity on the basis of mobile device which the user already adopted generally. According to large amount of data of cell-phone application and other mobile devices which are developing and popular, we can fulfill the mobile e-business on the basis of following several kinds of popular technology:

3.1 WAP

WAP is one of the key technologies of developing the mobile e-business. Through WAP, the cell-phone can conveniently and swiftly insert Internet whenever and wherever possible, really realizing the mobile e-business not restrained by time and region. WAP is a kind of communication protocol, it offers a set of open, unified technological platforms, the user can visit and obtain Internet or network information and various kinds of services inside the enterprise expressed with unified content very relaxed while using mobile devices. It defines one a suit of the interfaces of software and hardware. People can use cellular telephone receive and dispatch E-mail and surf Internet as using their computer. Besides, WAP has offered a kind of applicative development and running environment and it also support the most popular embedded operating system at present.

WAP can support the most wireless devices used at present, on the transmitting network, WAP support kinds of mobile networks now, for instance GSMs, CDMA, PHSS, etc. It also support the coming third generation of mobile communication systems. Now, many telecommunications companies have put out many kinds of WAP products already, including WAP gateway, instruments for applicative development and WAP cell-phone which offer the users the services such as online information, ticket ordering, mobile bank, games, shopping and so on.

3.2 GPRS

GPRS break through GSM offer thinking that circuit

exchange fix type by network, divide into groups and exchange mode introduce and get GSM network. It through increasing corresponding function entity and stand and go on transformations come and realize and divide into groups and exchange systematically to existing base only, Thus improve the utilization ratio of resources. GPRS set up and join fast, suitable for and convey little data quantity businesses or convey heavy data quantity businesses frequently. GPRS technology can offer the speed of data which is up to 115 Kbit/secs for moving the user, because GPRS exchanges because of dividing into groups, so users can keep the online forever.

3.3 The close quarter transmitting technology

Close quarter transmit technology that we say, mainly means infrared transmitting technology, "Bluetooth"(Bluetooth), 802.11, RFID (Frequency Identification of Radio) and remote control, etc. But relatively successful ones at present are the infrared ray technology and Bluetooth technology. Infrared ray, Bluetooth, compare to 802.11 technologies have some advantages: low costs, extensively use on the terminal stations like cell-phone, etc., correctly transmitting which reduces the danger of intercepting information, and it is difficult to be interfered by the external world in addition^[4]. Because of seeing the market prospects, Visa International, KDDI, UC Card, Toyota Finance to test the infrared ray payment business.

"Bluetooth" is as another kind of communication technology which is low cost, low power, small circle. It can make the equipment carry on communication in the short distance without the line cable. For example, we can use the cellular telephone to pay in front of the dispenser. It is the key technology of realizing the wireless electronic wallet. "Bluetooth" supports 64 kb/s transition and data transmission, whose transition distance is 10 - 100 ms with the network principle of "the principal and subordinate network".

3.4 other developing Technology

At present, the movable IP technology and the moving localization systematic technology grow vigorously, especially the moving localization system. It can offer the business based on position, for instance it can offer the information, such as local news, weather and hotel, etc. to the tourist and company staff who going out to handle official business. This technology will bring the enormous opportunity for the development of tourist industry, retail business, entertainments and catering trade, etc.

4. IMPLEMENTATION MODE AND THE WAY OF THE MOBILE E-BUSINESS

Because of its own characteristic, there will be very different for the small and medium-sized enterprise

using the technology of mobile e-business to realize the mobile operation when compared to the large-scale enterprise.

First of all, on the inputting of the hardware, large-scale enterprises has much fund to buy the large-scale SP platform, the hardware such as equipment of server that can offer mobile service, while it is impossible for small and medium-sized enterprise. But they have their own advantages: setting up one's own hardware system offering mobile service, conditionally choosing one or several ordinary servers, making which to be the special-purpose server with GSM, GPRS communication equipment offering mobile service again. It can save the enormous expenses which buys SP's platform, In addition, several small and medium-sized enterprises can buy a SP platform together. So long as they add their distinguishing codes, they can realize one's own business when the system running; Or they can rent SP's platform to test or expand the mobile business of the company, Thus in case of achieving the goal of using mobile service, the input and the cost are reduced.

Secondly, on the development of the mobile business, the large-scale company can organize its own specialized development group to carry on the technological development., while the small and medium-sized enterprise can set up the software system which suit one's own scale of company by buying the shaping software the developed interface and module to carry on the secondary development through the software company. Compared with systems of the large-scale company, this method has the characteristic with low cost, high flexibility, spread the operation advantages and gain the customer quickly.

The third, on commercial operation, because of the scale, the large-scale company will put into a lot of manpower, material resources, and carry on advertising to run and maintain it. Because the systems of small and medium-sized enterprise are based on jointly, rented SP platform and ordinary servers, the expenses of running and maintaining are relatively low. When offering the service, the flexibility should be stronger than that in the large-scale company, However, at the same time, small and medium-sized enterprise can apply the system to handle the daily official business reliably to in itself, which realize purpose on "one a suit of systems plant and use more" raising the systematic frequency of utilization to the maximum extent to improve the efficiency of market expanding and inside management of the enterprise.

The fourth, on proportion between inputting and outputting, because of the large input on the mobile business, large-scale companies have to obtain the income from heavy popularity which forming a large scale. Small and medium-sized enterprise has high flexibilities for investing less. If it offers the suitable services meeting the user's needs, it will get good income even it is small. Though this kind of income may be lower than incomes of the large-scale company, it is relatively considerable to its input.

5. THE SUMMARY

The development of the mobile e-business is a inexorable trend, though it experiences the twist and turn. How to seize the opportunity and how to open up a suitable soil to grow up for the small and medium-sized enterprise in our country is the key to be strong. According to the existed mobile application level, there are 4~5 kinds of relatively ripen technology which can be used into the real mobile e-business business. The question while the small and medium-sized enterprise taking the road to mobile e-business must be noticed are as the following: to summarize the characteristic and advantage of small and medium-sized enterprise, to take different strategy from the large-scale enterprise, to draw the successful experiences of the mobile e-businesses from domestic and international case, to pay attention to the flexibility of mode and implement, to reduce the cost of investment and accelerate to obtain incomes.

ACKNOWLEDGEMENT

ICEB2004 is supported by the National Natural Science Foundation of China.

REFERENCES

- [1] Jingyuan Wang, Quiet anabiosis of mobile e-business (移动商务悄然复苏), e.chinabyte.com, 2003/12/17
- [2] Guobin Fu, Analysis and discusses on tempting foreground of mobile e-business, (前景诱人的移动商务应用切入模式分析与研讨), e.chinabyte.com, 2003/12/09
- [3] Zheng jiang, Economic Daily, What about the informationization of small and medium-sized enterprises now, (中小企业信息化如今火候到几分), 2002/07/11
- [4] Shan Li, Challenge and Opportunity of mobile payment, (移动支付商机与挑战), 2003/05/29

Ubiquitous Computing Based Mobile Commerce : Toward the Ubiquitous Commerce

Sangjun Kim, Euiho Suh, Keedong Yoo

Dept. of Industrial Engineering, Pohang University of Science and Technology,
Pohang, 790-784, Rep of Korea
{nataiji, ehsuh, kdy}@postech.ac.kr

ABSTRACT

Mobile commerce is heading for advanced fourth generation (4G) mobile systems. However, rapid development of ubiquitous computing technology can implement and complement the 4G mobile systems. It enables anticipation that ubiquitous computing technology creates the new commerce, so called ubiquitous commerce. This paper analyzes the characteristics of mobile commerce service and ubiquitous computing service, and then extracts the core requirements for ubiquitous commerce. Moreover, this paper suggests technical requirements for the ubiquitous commerce. These requirements can lead the development of technology for the future and suggest the milestone that mobile commerce in present heads for.

Keywords : Mobile Commerce, Ubiquitous Computing Technology, Ubiquitous Commerce

1. INTRODUCTION

In the past few years, by technology advance, especially Information technology, mobile commerce has been marked as a new paradigm that is the next stage after electronic commerce and electronic business. However, appearing of new technology and increasing of customer needs promote development of highly advanced technology, and it presignifies another revolutionary paradigm shift. In this situation, it is valuable work that prospects the future about mobile commerce.

Generally, mobile commerce and mobile business are used reciprocally and mobile commerce is defined as any transaction with a monetary value - either direct or indirect - that is conducted over a wireless telecommunication network[3,19]. Mobile commerce is continuously prevailing and its market size is magnifying daily. Industry forecast statistics predict exponential growth in this sector

- International wireless data market is expected grow from 170 million to more than 1.3 billion subscribers between 2000-2004, equipping themselves with 1.5 billion wireless-capable handsets and other internet appliances by end 2004
(Source : Cahners Instat Group)
- Wireless internet users in the Asia-Pacific region alone will rise ten-fold from 20 to 216.3 million between 2000-2007 (Source : Strategis Group)

Growth of market and increase of customer needs lead the technological advance, and it accelerates translation of fourth generation(4G) mobile systems from third generation(3G) mobile systems[5,23]. 4G mobile systems are project to solve still-remaining problems of 3G systems as to the perspectives of network and device, these technologies share common characteristics with

ubiquitous computing technology that has been rapidly developing recently[6,9]. Furthermore, ubiquitous computing technology aims at highly advanced technology surpassing the existing mobile commerce technology[4]. It enables anticipation that ubiquitous computing technology not only evolves the existing mobile commerce but also creates the new commerce. In other words, we can predict new paradigm shift confidently that converts mobile commerce into a new commerce, so called ubiquitous commerce.

Based on this assumption, this paper will extract the essential requirements for ubiquitous commerce in order to progress from existing mobile commerce. So, in section 2, we analyze the characteristics of mobile commerce service and ubiquitous computing service by studying related work. Section 3 shows the essential requirements for ubiquitous commerce and the technical demands for achieving each requirement.

2. LITERATURE RIVIEW

2.1 Characteristics of mobile commerce service

There are many studies on mobile commerce. Table 1 shows the summary of these researches.

<Table 1> Studies of mobile commerce service characteristics

Author	Characteristics
Tsalgatidou And Pitoura (2001)	Terminal : small size keypad and screen, small resources including memory and disk capacity, reliance on the finite energy provided by batteries. (disadvantage) Wireless network : bandwidth restriction

	and network topology, variant bandwidth and burst traffic (disadvantage) Usability :location awareness, adaptivity, ubiquity, personalization, broadcasting (advantage)[19]
Camponove And Pigneur (2003)	Mobility : freedom of movement, ubiquity, localization, reachability, convenience, instant connectivity, personalization (advantage) device limitation, bandwidth limitation, expensive cost (disadvantage) Network effects : positive external effect Exclusive control over important asset : exclusive control over important assets exclusive control of MNO(Mobile Network Operator)[3]
Kim et al. (2003)	Efficiency : instant clearance, mobility, instant reaction Effectiveness : instant satisfaction of demand, usage of location information, personalized service[8]

Analyzing these characteristics synthetically, characteristics of mobile commerce service are derived from not characteristics of commerce itself, but technological features, especially handheld device and wireless network. Therefore, these characteristics can be represented in two criteria, mobility and connectivity, in terms of handheld device and wireless network[13].

Mobility means freedom of movement and is relevant to ‘place’concept. In mobility perspective, it brings a number of unique benefits of mobile commerce[3]. Ubiquity is the most obvious advantage of a wireless terminal. A mobile terminal can fulfill the need both for real-time information and for communication anywhere, independent of the user’s location[11]. Portability means that user can easily carry around mobile terminals. The smaller and lighter the devices are, the more portable they are[19]. Convenience is an attribute that characterizes a mobile terminal. Devices store data, are always at hand and are increasingly easy to use[11]. On the other hand, connectivity means overcoming of available time limitation and is relevant to ‘time’ concept. In the view of connectivity, mobile commerce serves instant connectable service. Instant connectivity means that it will be easier and faster to access information on the web anytime[11]. Combinations of mobility and connectivity provide rich characteristics for mobile commerce service. Reachability is important for many people who want to be in touch and be available to other people. With a mobile terminal a user can be contacted anywhere anytime. Localization of services and applications will add significant value to mobile devices. Knowing where the user is physically located at any particular moment will be the key to

offering relevant services that will drive users towards transacting on the network[11].

These characteristics are unique features of mobile commerce service. They are extracted by considering technological perspectives, such as device and network. Table 2 shows the summary of these characteristics.

<Table 2> Characteristics of mobile commerce service

Mobility	Connectivity
Portability Carrying the mobile device easily[19] Ubiquity The possibility of using services anywhere, independent of user’s location[11] Convenience Mobile devices are always at hand[11]	Instant connectivity Easy and fast access information on the web anytime[11]
Location awareness User’s location information can be exploit to offer location-based relevant services[11] Reachability User can be reached anywhere anytime[11]	

2.2 Characteristics of ubiquitous computing service

Mark Weiser stated that the most profound technologies are those that disappear and they weave themselves into the fabric of everyday life until they are indistinguishable from it[4]. Sakamura Ken in Japan emphasized ubiquitous network, and IBM suggested pervasive computing concept corresponding to ubiquitous computing. Like this, characteristics of ubiquitous computing technology are arguable and it’s difficult to define the characteristics of ubiquitous computing exactly. However, opinions that ubiquitous computing technology will change not only the existing commerce but also all around environments drastically are converged[4]. At this point of view, such a forecasting becomes a good evidence of possibility that ubiquitous computing technology will evolve the mobile commerce technology and create the new paradigm of commerce. Many researchers predict an advent of new commerce, which is called ubiquitous commerce. Generally, ubiquitous commerce is defined as convergence of electronic commerce and mobile commerce, etc.[6]. The realm of ubiquitous commerce, previously isolated and intentional actions become automatic and even invisible, where commercial interaction in every physical and digital sphere will be ever-present. In a ubiquitous commerce environment, objects are always on, always aware, and always proactive by ubiquitous computing technologies[14].

Since Mark Weiser suggested the characteristics of ubiquitous computing in 1991, many researchers have added and developed the characteristics of ubiquitous commerce services. Yoo et al.(2004) suggested 6 characteristics of ubiquitous commerce service; Embeddedness, Mobility, Nomadicity, Proactiveness, Invisibility and Portability[24]. Table 3 shows these characteristics.

<Table 3> Characteristics of ubiquitous computing service, Yoo et al.(2004)

Characteristics	Description
Embeddedness	small intelligent devices are embedded in the physical world and connected to the fixed and/or wireless network
Mobility	client devices must be operated under the mobile and flexible network infrastructure
Nomadicity	the system provides a rich set of computing and communication capabilities and services to nomads as they move from place to place in a way that is transparent, integrated, convenient, and adaptive
Proactiveness	the system needs to be self-triggered to capture a priori what its users want to increase the service quality
Invisibility	to be as unobtrusive as possible, enable the user to put as little data as possible
Portability	providing services with hands-free or at least one-handed light device

2.3 Technical approach of 4G mobile systems and ubiquitous computing technology

As we have referred before, mobile commerce aims at fourth generation(4G) mobile systems[5,23]. Generally, 4G mobile systems should provide various multimedia services, ubiquitous services and extended service through seamless global roaming[7]. According to ITU-R WP8F's vision and objective of 4G, 4G mobile systems have characteristics of ubiquitous & seamless connection, high data rate, openness, and network convergence. 4G mobile systems have to solve these problems concretely [5,7].

- Average 200 Mbps high data rate
- More 10 times increasing capacity than 3G
- 3~6 GHz broad bandwidth
- QoS problem

- Mobility management

Table 4 shows hot technology issues to solve existing problems of 3G mobile systems.

<Table 4> Core technology for 4G mobile systems

Technologies	Description
MIMO (Multi-Input, Multi-Output)	- Both the transmitter and receiver use multiple antenna elements, or antenna arrays - Help mitigate channel impairments and help reduce interference, thus yielding better error rate performance to support an increased number of users
UWB (Ultra Wide Band)	- Transmit and receive over a wide range of frequencies supporting a variety of radio standards used around the world - Can be used in a broad array of applications ranging from voice and data communication to high - performance radar systems
SDR (Software Defined Radio)	- Performed by software residing in high-speed digital processors - The radios can consequently be easily reprogrammed to transmit and receive over a wide range of frequencies supporting a variety of radio standards used around the world

Ubiquitous computing technology can solve and complement these problems that 4G mobile systems pursue. Table 5 shows element technology of ubiquitous computing.

<Table 5> Technology classification table of ubiquitous network project in Japan

Technical type	Technology
Ubiquitous system technology	Flexible personalized system tech. Mobility control tech. Advanced location awareness tech. Profile portability tech. Advanced sensing system tech. Data-grid tech. Ubiquitous address operation management system tech. Architecture tech. Real-time OS tech.
Network technology	Seamless network access tech. Photonic network tech. Full IPv6 network tech. High capacity wireless tech. Zero administration tech. Network QoS tech. Distributed network tech.

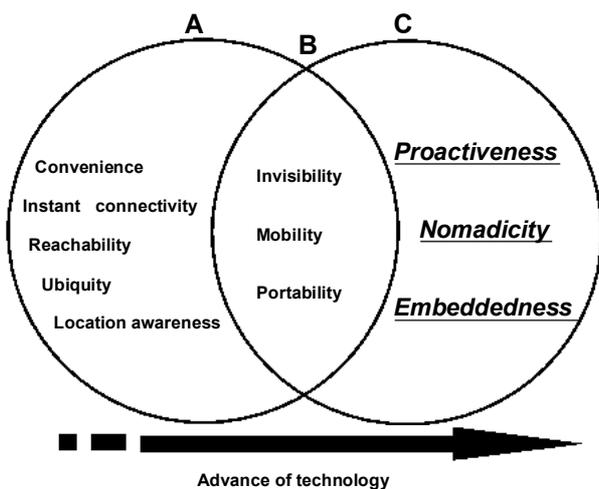
Application technology	Intelligent contents tech. Transcoding tech. U-agent tech. Optimistic and inference engine tech.
Appliance technology	MEMS Advanced electric power tech. Micro interface tech.
Platform technology	IC card certification tech. Personal certification tech. DRM tech. Optimistic security system tech.

3. ESSENTIAL REQUIREMENT FOR UBIQUITOUS COMMERCE

3.1 Technical relativity

Considering the aims of 4G mobile systems and the technology of ubiquitous computing, Ubiquitous computing technology can not only solve the limitation of 3G mobile systems but also includes core technology for 4G mobile systems. For example data rate and capacity problems can be solved by network technology of ubiquitous computing. Additionally, fatal limitation of mobile commerce, especially small keypad and screen or computing capacity, can be overcome by using platform and appliance technology. Moreover, the core technologies for 4G mobile systems, such as MIMO, UWB and SDR, are included in ubiquitous computing technology ; ubiquitous system technology, network technology and application technology.

Then, what is the requirements for ubiquitous commerce? We can extract these requirements by comparing characteristics of mobile commerce service and ubiquitous commerce service. Figure 1 shows relativity of each characteristics.



<Figure 1> Ubiquitous commerce requirements diagram

Part A is the characteristics of mobile commerce service only. These characteristics are already developed. Part B is the shared-region of mobile and ubiquitous commerce and these characteristics are under development. Part C is the unique region of ubiquitous service only. This region is not developed yet and is an important part we have to focus on. Therefore, elements in part C become essential requirements for ubiquitous commerce.

3.2 Nomadicity

Nomadicity means advanced network technology. The system needs to provide a rich set of computing and communication capabilities and services to nomads as they move from place to place in a way that is transparent, integrated, convenient and adaptive[24]. To accomplish this requirement, following technical demands must be developed.

- Ubiquitous system technology

Real-time O/S technology, Mobility control technology, Advanced location awareness technology

- Network technology

Seamless network access technology, Full Pv6 network technology, Network QoS technology

3.3 Embeddedness

Small intelligent devices are embedded in the physical world and connected to the fixed and/or wireless network[24]. In other words, small device disappeared and user can be freed from the device. Embeddedness can create new business model. Health care toilet of Matsushita Electric Industrial Co. in Japan is a good example. Following technical demands must be developed.

<Table 6> Technical demands of each requirement

	Nomadicity	Embeddedness	Proactiveness
Ubiquitous system technology	Real-time O/S tech. Mobility control tech. Advanced location awareness tech.	Advanced sensing system tech.	Advanced sensing system tech.
Network technology	Seamless network access tech. Full IPv6 network tech. Network QoS tech.		Network QoS tech.
Application technology			U-agent tech. Optimistic inference engine tech.
Appliance technology		MEMS Advanced electric power tech.	Micro interface tech.

•**Ubiquitous system technology**

Micro sensing system technology

•**Appliance technology**

MEMS(Micro Electro Mechanical System),
Advanced electric power technology

3.4 Proactiveness

The system needs to be self-triggered to capture a priori what its users want to increase the service quality[24]. Proactiveness has the same meaning as context awareness. Context is defined as any information that can be used to characterize the situation of an entity[16]. Context awareness is a very important concept of ubiquitous computing. Without context awareness, there's no ubiquitous commerce[15]. Yoo et al.(2004) emphasized the importance of context awareness[24]. Proactiveness needs mixed-application of various advanced technology because of technical complexity.

•**Ubiquitous system technology**

Advanced sensing system technology

•**Network technology**

Network QoS technology

•**Appliance technology**

Micro interface technology

•**Application technology**

U-agent technology, Optimistic and inference engine technology

4. CONCLUSION AND FUTUREWORK

In present, mobile commerce is heading for advanced 4G mobile systems. But specific technical element of 4G mobile systems can be implemented and complemented by ubiquitous computing technology that is under development. Furthermore, ubiquitous computing technology will create new paradigm of commerce, so called ubiquitous commerce. Based on this assumption, this paper analyzed the characteristics of mobile commerce service and ubiquitous commerce service and extracted the requirements for ubiquitous commerce, which are Embeddedness, Nomadicity and Proactiveness. Finally, we suggested core technical demands.

This paper depends on uncertain future technology heavily. This is a fatal limitation of this paper. Therefore, we must study more cases of mobile commerce and ubiquitous computing technology, and for credibility of the requirements, there is a need to carry out case study and apply reliable methodology such as survey and factor analysis. In spite of this limitation, requirements extracted in this paper can suggest milestone that existing mobile commerce goes for, and we can expect the development of technology.

REFERENCES

- [1] Barkhuus, L.(2002), Ubiquitous Computing: Transparency in Context-Aware Mobile Computing, Fourth international conference on ubiquitous computing, Sweden
- [2] Barnes, S.(2002), The mobile commerce value chain: analysis and future developments, International Journal of Information Management 22 91-108
- [3] Camponovo, G. and Pigneur, Y.(2003), Business model analysis applied to mobile business, 5th International Conference on Enterprise Information Systems, France
- [4] Fano, A. and Gershman, A.(2002), The Future of Business Services in the Age of Ubiquitous Computing, Communication of the ACM, 45(12), 83-87
- [5] Gazis V., Houssos N., Alonistioti A., Merakos L.(2002), Evolving perspectives of 4th generation mobile communication systems, The 13TH IEEE International symposium on personal, indoor and mobile radio communications, September, Lisbon
- [6] Gershman, A.(2002), Ubiquitous Commerce – Always On, Always Aware, Always Pro-active, The International Symposium on Applications and the Internet, Nara, Japan
- [7] Ibrahim, J.(2002), 4G Features, Bechtel Telecommunications Technical Journal 1(1) 11-14
- [8] Kim S., Han C., Kim S. and Park Y.(2003), Major characteristics and function of mobile business, Proceeding of Korean operations research and management science society 2003, 276-279, Pohang, Korea

- [9] Kleinrock L.(2003), An internet vision : the invisible global infrastructure, *Ad Hoc Networks* 1 3-11
- [10] Kumar, S. and Zahn, C.(2003), Mobile communications: evolution and impact on business Operations, *Technovation* 23 515–520
- [11] Muller-Veerse F.(1999), Mobile commerce report, Durlacher, London
- [12] Nakajima H.(2002), Marketing Strategy in the Era of Ubiquitous Networks, *NRI papers*, 44
- [13] Park C., Yang H. and Ahn J(2003), Developing a Mobile-Business Framework Considering Ubiquitous Computing Environment, *information systems review* 5(2) 37-51, *Korean Journal*
- [14] Park D.(2003), Co-evolution in uCommerce: Emerging Business and Technologies, *Telecommunications Review* 13(1) 48-56.
- [15] Pitkänen, O.(2003), Legal challenges to Ubicommerce, Fifth international conference on ubiquitous computing, seattle
- [16] Prekop, P. and Burnett, M(2003), Activities, context and ubiquitous computing, *Computer Communications* 26 1168–1176
- [17] Sarker, S. and Wells, J.(2003), Understanding Mobile handheld device : use and adoption, *Communication of the ACM*, 46(12) 35-40
- [18] Tarasewich, P.(2003), Designing Mobile Commerce Applications, *Communications of the ACM*, 46(12), 57-60
- [19] Tsalgatidou, A. and Pitoura, E.(2001), Business models and transaction in mobile electronic commerce : Requirements and properties , *computer networks* 37 221-236
- [20] Varshney, U. and Jain, R.(2001), Issues in Emerging 4G Wireless Networks, *Computer June* 94-96
- [21] Venkatesh, V., Ramesh, V. and Massey, A.(2003), Understanding usability in mobile commerce, *Communications of the ACM* 46(12) 53-56
- [22] Weiser, M. and Brown, J.(1996), The coming age of calm technology, *Xerox PARC*
- [23] Willie W.(2003), 4G Mobile Research in Asia, *IEEE Communication Magazine* March 92-95
- [24] Yoo K., Kwon O., Suh E.(2004), UbiDSS : a proactive intelligent decision support system as an expert system deploying ubiquitous computing technologies, *Expert systems with applications*, article in press

User-Centered Guidelines for Design of Mobile Applications

Xiaowen Fang, Susy Chan, Jacek Brzezinski, Shuang Xu, Jean Lam

School of Computer Science, Telecommunications, and Information Systems, DePaul University,
243 South Wabash Avenue, Chicago, IL 60604, USA
{xfang, schan, jbrzezinski, sxu}@cti.depaul.edu, jeanlam@us.ibm.com

ABSTRACT

In this study, interviews were conducted to derive user-centered guidelines for the interface design of mobile applications. These guidelines cover general design issues in input, display, navigation, and feedback. They address five out of seven elements that describe effective mobile commerce interface design proposed by Lee and Benbasat [7]. Compared to guidelines obtained in prior studies, the guidelines derived from this study are user-centered and can be applied to multiple form factors. The comparison between e-commerce and mobile commerce design guidelines suggests that the mobile context plays the most important role in designing mobile commerce interfaces.

Keywords: mobile commerce, design guidelines, usability, human-computer interaction

1. INTRODUCTION

The convergence of mobile Internet and wireless communication technology has promised users anytime, anywhere access to information for work and personal communications. However, many technology constraints such as small screen size hinder such access through a mobile device. In order to provide truly ubiquitous access to information, design guidelines addressing potential usability problems on mobile devices are essential.

Usability studies on mobile applications have focused on addressing the interface design constraints imposed by a bandwidth limitation and small screen size of handheld devices. Along this line, researchers have proposed several design guidelines for Wireless Application Protocol (WAP) phone applications. Ramsay and Nielsen [8] suggested that many WAP usability problems echoed problems identified during the early stages of Web site development for desktop computers, and could be alleviated by applying good user interface design. Such design guidelines for WAP applications may include: (1) short links, (2) backward navigation on every card, (3) a minimal level of menu hierarchy, and (4) headlines for each card [5]. Similar design guidelines, validated in a separate usability study of WAP phones [2], include: (1) direct, simple access to focused valuable content, (2) simple hierarchies, (3) reducing the amount of vertical scrolling, and (4) reducing the number of keystrokes. These studies focused solely on WAP phones.

Diverse form factors offer different functionalities and have separate interface requirements. Chan et al. [4] systematically reviewed 10 wireless Web sites – ranging from travel, financial services, retail, news, and Internet portals – across three form factors: wireless Palm, WAP phones, and Pocket PCs. They found that user tasks for the wireless sites were designed with steps similar to the wired e-commerce sites and geared primarily for

experienced users. Many usability problems, such as long downloads and broken connections, information overload, and excessive horizontal and vertical scrolling, are common to all three form factors. Based on this study, interface design flaws are platform independent, but the more limitations are imposed on the form factors, the more acute the design problems become. These researchers recommend eight guidelines: (1) avoid scrolling, (2) use a flat hierarchy, (3) design a navigation system consistent with a regular Web browser, (4) include a back button, (5) provide a history list, (6) provide indication of signal strength, (7) reduce user's memory load, and (8) limit the search scope to improve search efficiency. Furthermore, Chan and Fang [3] indicated that limited bandwidth and multiple form factors posed constraints for user interface design in terms of the amount and format of content presentation, navigation, and site structure.

Lee and Benbasat [7] proposed a reference framework for designing m-commerce interfaces. This framework consists of seven design elements: context, content, community, customization, communication, connection, and commerce. Venkatesh, Ramesh, and Massey [10] assessed usability guidelines for Web and wireless sites and found that ease of use was significantly more important in wireless contexts. Tarasewich [9] has also suggested that context plays an important role in the design of mobile commerce applications. Furthermore, design guidelines and usability methods that work with wired systems do not necessarily work with mobile systems.

However, the design guidelines developed so far have two major shortcomings: 1) Some guidelines may be only applicable to WAP applications; and 2) Some guidelines were derived from expert opinions without real user input. In order to develop user driven design guidelines that are applicable to a broad range of form factors, a study was undertaken as discussed below. In this study, participants were asked to use one of three

types of handheld devices: WAP phone, Palm Pilot, and Pocket PC. Their experience and opinions were elicited through interviews.

2. METHOD

A wireless web site that provided the function of making an advising appointment was developed on three platforms: Palm, WAP phone and Pocket PC. Similar to the regular web site of School of Computer Science, Telecommunications and Information Systems at DePaul University, these wireless versions of a desktop application allow students to make advising appointments with faculty members. Fig. 1 shows sample screen shots of this Web site on the three platforms. Thirty-seven students from DePaul University participated in the experiment. A pre-experiment questionnaire was used to collect background information on the participants. Questions were asked regarding their familiarity with the functionality of the regular web site and their experience with WAP phone, Pocket PC, Palm, and other wireless devices. Among these participants, some had used handheld devices before, while others had not. Most of them had used the advising appointment application on the regular web site before the experiment.



Fig. 1. Sample screens of the experiment Web sites

Nineteen participants were assigned to perform tasks on Palm, nine participants on WAP phone, and nine participants on Pocket PC. Participants were asked to make appointments with their faculty advisors using the handheld device. While the participant was performing the task, the experimenter sat aside, observed the participant's performance, and took notes about any unusual events such as mistakes, comments, and complaints. Each participant's online activities were also recorded into a log file by a server-side program. Upon completion of the task, each participant answered questions during an exit interview as to what they liked or disliked about using the handheld device and about the appointment application.

During the interview, if the participant used any ambiguous terms, follow-up questions were asked for clarification. Each interview took about 45 minutes. All

interviews were tape-recorded and transcribed. Based on the transcribed comments, we distilled findings that were independent of the advising application through the follow process:

- Useful comments were extracted from the transcripts and rephrased to reflect similar comments based on authors' professional judgment.
- Each new comment was assigned a unique number and marked with its recurring frequency in the participant's commentary.
- A list of comments for each participant was created.
- Another list was constructed to combine comments from all the participants. Similar or identical comments were consolidated and listed as one.
- Comments were further summarized into categories.

From these findings, we derived user-centered design guidelines as discussed below.

3. USER-CENTERED DESIGN GUIDELINES

Table 1 presents 18 most important design guidelines derived from this study. These design guidelines address users' concerns from the following four aspects of a mobile application: input, display, navigation, and feedback. Because these design guidelines reflect users' response to mobile applications, the findings could be extremely valuable to developers of mobile applications.

The reference framework for mobile commerce interfaces proposed by Lee and Benbasat [7] was used to analyze these design guidelines. As shown in Table 1, the design guidelines from our study address five of the seven design elements for mobile commerce interface design: context, content, customization, communication, and commerce.

Context captures how Web sites are developed, consisting of functionality and aesthetics [7]. Design guidelines concerning navigation of a wireless Web site address this element. Given the mobile setting, navigation functions must be designed to accommodate the context.

Content focuses on what a site presents, comprising the offering, appeal, multimedia mix, and content type [7]. This study indicates that large amount of graphics is not suitable due to the small screen of handheld devices and security information must be presented.

Customization refers to a site's ability to tailor itself or to be tailored by users [7]. Customization reduces information load by filtering unnecessary information. This is consistent with the results from this study. One of the guidelines states that short and concise information display is preferred.

Table 1. User-centered design guidelines for mobile applications

Category	Design Guidelines	Interface Design Elements
Input	<ul style="list-style-type: none"> • Input on handheld devices must be minimized • Graffiti is preferred by users. It is faster but less accurate compared to soft keyboard. 	Communication
Display	<ul style="list-style-type: none"> • Short and concise displays are preferred on handheld devices. A large amount of graphic and audio/video information is not suitable for handheld devices. • Features indicating security protection must be prominent on handheld devices. 	Content / Customization
	<ul style="list-style-type: none"> • Products that do not have to be graphically represented such as books and audio CDs are more suitable for online shopping on handheld devices. 	Commerce
	<ul style="list-style-type: none"> • Widely-accepted short names are appropriate for the items in a drop-down menu. 	Content
Navigation	<ul style="list-style-type: none"> • Wireless Web sites must be consistent with regular Web sites. • Exit from an application on handheld devices must be visible. • A hierarchical information structure is helpful and preferred. • A history list is important in the navigation process. • Buttons and menus must be presented on the top of a Web page on handheld devices. • Icons with descriptive and accurate labels are preferred. • Hyperlinks and buttons must be prominent. • Scrolling should be avoided if possible. If scrolling cannot be avoided, use “page-up” and “page-down” type scrolling. • Buttons must be clearly separated from each other. • Access to action buttons in a form should not be obscured by scrolling. • Scrolling in a drop-down menu must be avoided. Items in a drop-down menu should be sorted. 	Context
Feedback	<ul style="list-style-type: none"> • For successful submissions or tasks involving multiple steps (pages), confirmation is necessary. 	Communication

Communication is defined as dialogue between sites and users: broadcast, interactive, and hybrid [7]. The design guidelines from our study indicate that during the interaction between sites and users, information input on handheld devices must be minimized and feedback must be provided.

Commerce is concerned with interfaces related to sales of goods and product services [7]. It was found in our study that features indicating security protection must be prominent. It was also found that the so called “low touch” products like books or audio CDs do not require graphic display and are more suitable for online shopping using handheld devices.

Some of the most important findings are further discussed in the following sections.

3.1 Text entry

Text entry on the handheld devices is inconvenient due to lack of efficient input methods. About 90% of the participants brought up this issue. The findings of our study suggest:

- Text entry or typing must be minimized. Users don't like to enter long texts regardless of the input methods: graffiti, soft keyboard, or keypad. No matter which input method is used, it usually takes

significantly more effort to type texts on handheld devices than on regular PCs.

- Graffiti is preferred by experienced users because it is faster than soft keyboard. However, it takes users more time to learn graffiti and graffiti is usually less accurate than soft keyboard. Novice users prefer soft keyboard.

3.2 Information display

Information overloading and over-crowded screens are the problems most commonly shared by regular Web applications and mobile applications. However, these problems become more severe for mobile applications because of two factors: the small screen size of handheld devices and the user's mobility. Our study has identified the following unique requirements:

- Short and concise output is preferred. To minimize excessive amount of text information, widely-accepted short names can be used in the display.
- In general, users like graphical icons and buttons but a large amount of graphical and audio/video information is not suitable for handheld devices.
- Scrolling should be avoided if possible. Scrolling is significantly more difficult on handheld devices than on PC computers. Horizontal scrolling is more confusing than vertical scrolling.

- Action buttons or links such as “Submit” and “Search” should not be separated from the input text boxes. Due to the small screen, users are often confused by the separation when trying to figure out how to submit the form.
- Products that do not require graphic display such as books and CDs are more suitable for online shopping on handheld devices. For this type of products, users are not concerned about the graphical details, which cannot be displayed clearly on handheld devices.

3.3 Security

Security has been a major concern in e-commerce. It becomes an even more sensitive issue for mobile applications due to doubts about the security and reliability of wireless connections. About 40% of the participants of this study expressed their concern regarding security with mobile applications. They frequently mentioned that they were not willing to commit online transactions on a handheld device because they worried about the security and reliability of the wireless connection. For mobile applications that involve any online transaction, security assurance is important.

3.3 Tasks

Tasks that can be performed on handheld devices are different from what can be executed on PC computers. As suggested by Ankar and D’Incau [1], mobile applications must address the following five most common mobile values: (1) time-critical needs and arrangement; (2) spontaneous needs and decisions, such as auction, email, and news; (3) entertainment needs; (4) efficiency needs and ambitions; and (5) mobility related needs. In a separate study, Xu, Fang, Chan, and Brzezinski [11] identified five factors contributing to user’s preference of tasks to be performed on handheld devices: perceived usefulness, perceived ease of use, perceived playfulness, complexity, and perceived security. The following users’ responses are consistent with these two studies:

- Users would like to perform tasks that will meet their mobile needs such as sending/receiving emails, and checking flight status.
- Users would like to see an easy and friendly user interface with considerations of all the constraints imposed by handheld devices and mobility.
- Users would like to use handheld devices to entertain themselves in their free time.
- Users would like to perform simple tasks.
- Users are concerned about the security of online transactions.

4. DISCUSSION

Usability for regular e-commerce Web sites has received a lot of attention in both academia and industry. In this

section, we will compare design guidelines for e-commerce and mobile commerce and discuss what is most important to mobile commerce users.

Fang and Salvendy [6] conducted a series of interviews to derive customer-driven rules for the design of e-commerce Web sites. Their study captured a broad range of e-commerce Web sites and tasks, and developed the following design guidelines:

- Web page should be clean and not cluttered with text and graphics.
- The width of a page should be less than the width of the browser window to avoid horizontal scrolling.
- Text on the links or buttons should be self-explained and descriptive.
- When linking to another product related Web site, link to the exact product page instead of the homepage of that site.
- Categorize products in a way that is meaningful to regular customers.
- The depth of the categories should be no more than 3.
- Present accurate, consistent, and detailed descriptions of products.
- Provide accurate and full pictures of products.
- Present the size of products in a measurable and comparable way.
- Present the inventory information of a product in the beginning.
- Present products in a table with enough information to make a purchasing decision such as prices and features for easy comparison.
- Present related charges up front and in an accurate way.
- Same products should be presented in the same page, same position.
- Products shouldn’t be removed from the page because of out-of-stock.
- In the shopping cart page, provide a link that directs the customer back to the page he/she left for continuing the shopping.
- Only ask for necessary and meaning information such as name and address. No marketing questions.
- Allow customers to browse the site without logging in.
- Provide a 1-800 number for customers to call.
- Clearly state the return policy in a prominent place.

A comparison between design guidelines for e-commerce sites and mobile commerce sites reveals the following:

- Both e-commerce and mobile commerce sites share some common design guidelines such as avoiding horizontal scrolling and labeling buttons and links clearly.

- Most e-commerce design guidelines are concerned with offering specific and accurate information.
- Most mobile commerce design guidelines are concerned with design constraints imposed by mobile context such as device limitations and user's mobility. Mobile users can only input and view very limited amount of information on handheld devices under mobile conditions. Design of mobile applications must satisfy user's needs and desires.

Findings from the current study are consistent with prior research findings. Mobile context was identified as the most important factor in designing mobile commerce interfaces in two separate studies [9] [10].

5. CONCLUSIONS

In this study, user-centered guidelines for design of mobile applications were developed. These guidelines cover general design issues for input, display, navigation, and feedback. They address five out of seven elements for effective mobile commerce interface design proposed by Lee and Benbasat [7]. Compared to guidelines derived from prior studies, these guidelines are user-centered and can be apply to multiple form factors. The comparison between e-commerce and mobile commerce design guidelines suggest that mobile context plays the most important role in designing mobile commerce interfaces. However, more studies are needed in the future to address the technical constraints imposed on interface design for mobile devices.

REFERENCES

- [1] Anckar, B., and D'Incau, D. Value creation in mobile commerce: Findings from a consumer survey. *Journal of Information Technology Theory & Application*, Vol. 4, pp43-64, 2002.
- [2] Buchanan, G., Farrant, S., Jones, M., Thimbleby, H., Marsden, G., & Pazzani, M. Improving mobile Internet usability. *Proceedings of the tenth International World Wide Web Conference on World Wide Web*. New York, NY: ACM. pp673-680, 2001.
- [3] Chan, S., & Fang, X. Mobile commerce and usability. In K. Siau & E. Lim (Eds), *Advances in Mobile Commerce Technologies*. Hershey, PA: Idea Group. pp235-257, 2003.
- [4] Chan, S., Fang, X., Brzezinski, J., Zhou, Y., Xu, S., and Lam, J. Usability for mobile commerce across multiple form factors. *Journal of Electronic Commerce Research*, Vol. 3, 187-199, 2002.
- [5] Colafigli, C., P. Inverard, and R. Martriccian. InfoParco: An Experience in Designing an Information System Accessible through WEB and WAP Interfaces. *Proceedings of the 34th Hawaii International Conference on System Science*, Los Alamitos, CA, IEEE Computer Society Press, 2001.
- [6] Fang, X., and Salvendy, G. Customer-centered rules for design of E-Commerce Web sites. *Communications of ACM*, Vol. 46, No. 12, pp332-336, 2003.
- [7] Lee, Y., and Benbasat, I. Interface design for mobile commerce. *Communications of ACM*, Vol. 46, No. 12, pp49-52, 2003.
- [8] Ramsay, M., and Nielsen, J. WAP usability: déjà vu: 1994 all over again. Unpublished report. Nielsen Norman Group, 2000.
- [9] Tarasewich, P. Designing mobile commerce applications. *Communications of ACM*, Vol. 46, No. 12, pp57-60, 2003.
- [10] Venkatesh, V.; Ramesh, V.; and Massey, A. Understanding usability in mobile commerce. *Communications of ACM*, Vol. 46, No.12, pp53-56, 2003
- [11] Xu, S., Fang, X., Chan, S., and Brzezinski, J. What tasks are suitable for handheld devices ? In *Proceedings of the 10th International Conference on Human-Computer Interaction*, Crete, Greece. Pp333-337, 2003.

Using System Dynamics to Model Mobile Commerce Diffusion

Wenqing Wang¹, France Cheong²

School of Business Information Technology, Faculty of Business, RMIT University

Level 17, 239 Bourke Street, Melbourne, Vic 3000, Australia

¹wenqing.wang@rmit.edu.au, ²france.cheong@rmit.edu.au

ABSTRACT

Mobile commerce is increasingly perceived as a powerful lever to foster growth of local economy and an opportunity to generate revenue. However, both the literature and empirical evidence have revealed that there exist perils hidden in decision-making by mobile commerce advocates. A system dynamics (SD) approach is used in this article to analyze how factors will influence the process of mobile commerce diffusion. In this paper, mobile commerce market formation is studied as a social economic system, which is dynamic and whose states change continuously. The model-building process incorporates a wide variety of sources of knowledge ranging from system thinking, through international marketing and management science, to system dynamics. The resulting simulation models are developed upon the mental models of a group of experts and available information from written documents, web forum, and online workshops. A simulation model is empirically established and simulated for the Chinese market. The model development and related analysis provides a typical example in using system dynamics to understand the process of mobile commerce diffusion and decision-making.

Keywords: mobile commerce, electronic commerce, system dynamics, innovation diffusion, simulation, modeling

1. INTRODUCTION

The fact that m-commerce is developed and rooted in Western countries makes m-commerce implementations more Western-oriented. Most of the m-commerce services are created to meet the requirements of the pre-existing infrastructures and environments of developed economies. This situation will make developed markets and emerging markets totally different in evaluating the market potential. Is there a generic model that is globally applicable in simulating each of the existing developed or emerging markets? System thinking and system dynamics are relevant methods in achieving this purpose. In order to accurately reflect m-commerce market, it is necessary to put mobile commerce diffusion under system thinking [1-3], in which the mobile commerce market is regarded as a complex system and the relationship between its existing structure and the behavior will be explored in a dynamic situation. Efforts in this area are likely to help entrepreneurs in learning the mobile commerce market and understanding opportunities and pitfalls related to m-commerce decision-making.

In order to explore the mobile commerce market, we developed a system-learning framework. In the real world, there are two categories of mobile commerce markets. One is called the existing mobile commerce market, or similar market like mobile communication market, which already exists in some developed countries. The application of mobile communication service and mobile commerce service in these countries provides important empirical evidence to explore new emerging markets. Another category is the emerging mobile commerce market in developing countries. Because mobile commerce in emerging area is still in its

infancy, the market is in a dynamic state. The structure is still unknown. Large-scale experiments about the system are inevitable if there is no other method available. In fact, any controlled experiments about a countrywide dynamic market may turn out to be expensive and unrealistic. Whereas developing virtual world model to simulate the real world and doing experiments on the computer-based model is more cost effective and operative. As Sterman [4] points out, a virtual world model provides low-cost experimentation for learning complex system without the limitation of time and space. This paper is aimed at introducing the modeling of the virtual world.

System thinking and system dynamics are effective tools to explore the world. Information collected from the real world will be analyzed and synthesized using system thinking [2, 5], system dynamics [6-11] and innovation diffusion theories [12, 13]. Group experts' opinion is solicited to enforce the modeling process. A model is finally developed with main feedback loops and logical and mathematical relationship coupling the key variables. Since the research is still in progress, the empirically developed model will be further tested using the data collected from existing markets. After testing, the model will be applied to other existing and emerging markets.

Our research framework is divided into the four stages: stage 1, analysis of mobile telecommunications industry, management science, control theory, computer simulation and Delphi surveys, aimed to establish theory and to build qualitative System Dynamics (SD) models; stage 2, development of a generic SD simulation model based on the insights from stage 1; stage 3, collection of several secondary data and primary data in existing

market, aimed to calibrate, validate the model developed in the previous stage; stage 4, scenario analysis and corresponding strategy design in any target market.

Up to now, stages 1 to stage 2 have been finished and a SD model has been built. This article aims to outline the findings generated so far from the research. In order to provide a clear outline for the paper, the main critical loop and stock and flow analysis of m-commerce diffusion is initially outlined. Then, modeling analysis is presented. Such an analysis aims to depict the main forces controlling the flow rates associated with local policies and other participants' strategies [5-7]. Finally, the main research findings and implications for further research are summarized.

2. MODEL DEVELOPMENT

Mobile commerce diffusion can be considered as a technological innovation. The diffusion and adoption of news ideals and new products often follows S-shaped growth patterns [4]. The problem is to identify what the positive feedback that generate the initial exponential growth of a successful mobile commerce diffusion are and what the negative feedback that limits the growth are and, to what degree, any endogenous or exogenous variables will influence the growth both in magnitude and speed of reaching an equilibrium.

2.1 Main Causal Loop and Stock Flow

The system boundary of the research refers to any mobile commerce market covering a geographical country or region, in which a certain number of populations reside. The number of people in the population fluctuates with the variation of net increase rate (birth rate minus death rate), which is controlled by local population policy through adjusting birth rate. In the population pool, a proportion of the population has the Capability to use a mobile terminal (hereafter refers to mobile phone) because their income and knowledge make them eligible to consume the mobile phone service although they may have not taken the service. This part of the population is called capable population. From the capable population, some of them, who have adopted the mobile phone services, are called mobile terminal users. In the emerging mobile commerce market, the 3G licenses are not issued yet (3G refers to the third generation of telecommunication networks). It means up to now there is no actual mobile commerce user but the mobile terminal users are the potential ones. As long as there is a demand from the market, mobile terminal manufacturers will replace ordinary mobile terminals with mobile terminals with mobile commerce capability, over which commerce services will be conducted. Then the mobile terminal users will gradually consume the mobile commerce services and become actual mobile commerce users.

Figures 1 and 2 illustrate the causal loop diagram and stock flow diagram identified for the main system. To visualize it, imagine the whole market system is like a water irrigation system with the population as the source at the top; the flow starts at Potential Capable Population, passing through Potential Mobile Terminal Users, and returns back to Potential Capable Population. The process forms a typical closed causal loop. As illustrated in Figure 1, the Actual Mobile Commerce Users come from Actual Mobile Terminal Users, the Actual Mobile Terminal Users from the Actual Capable Population from the population reservoir in the origin. The clockwise closed loop starting from the Potential Capable Population through the Potential Mobile Terminal Users to the Potential Mobile Commerce Users illustrates the evolution, development and formation of mobile commerce diffusion. It is along this critical process that the research starts to explore the system structure and discover the market behavior over time both in the short term and in the long term.

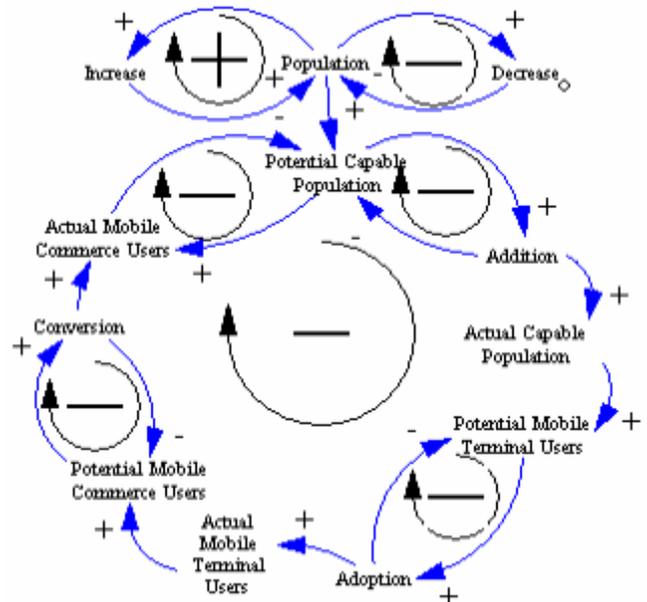


Figure 1 Causal loop diagrams in main system

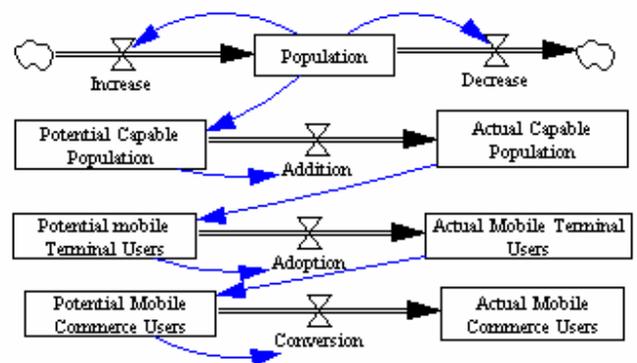


Figure 2: Stock and flow diagram in main system

The whole system includes one big closed feedback loop and four semi closed feedback loops (excluding the population loop). There are two fundamental loops that

affect the number of Mobile Commerce Users. One is the clockwise loop from the Potential Capable Population through the Potential Mobile Terminal users and the Potential Mobile Commerce Users back to the Potential Capable Population. In this loop, starting from the Potential Capable Population in a clockwise direction, an increase in value of each variable will result in the increase in value of the next variable except that an increase in Actual Mobile Commerce Users will reduce the level of Potential Capable Population. Thus, the loop constitutes a negative loop (in Figure 1, B, short for balancing, is used to denote a negative loop; R, short for reinforcing, is used to indicate a positive loop). Another important loop is the semi loop directly linking the Potential Capable Population and the Actual Mobile Commerce Users. An increase of the Potential Capable Population will increase the value of the Actual Mobile Commerce Users, but, on the contrary, an increase of the Actual Mobile Commerce Users will decrease the value of the Potential Capable Population. This semi loop is also a typical negative loop. The combination of these two fundamental loops determines the state of the Actual Mobile Commerce Users. With these loops combined with variables, a variety of behavior patterns are possible [7]. In this research, the big negative loop, as well as any semi negative loop, generates flows that add to each level downstream to the Actual Mobile Commerce Users and tries to generate goal-seeking behavior. It is the goal and its development that the research tries to seek.

Another important loop (stock) that is the beginning of the main stream at the top of Figures 1 and 2 is related with the population state. The population is normally regarded as an exogenous factor and defined as a constant. In this research, the population is presumably regarded as an endogenous factor because population policy is widely manipulated as a tool to change long-term conditions in most of the populous developing countries. For example, family planning is used as a leverage to control the population birth rate. The population in the system is the only source of the Potential Capable Population, which is the starting point of the flow generated in the main loop, so the fluctuation in the population will inevitably vary the flow in the main loop.

2.2. Simulation Model

A generic model, developed using the PowerSim software package, focuses on the adoption and diffusion of mobile commerce service. The factors that influence the change rate in each level are derived from the related theory and experts' opinion. This section examines the assumptions and equations contained in each level. In the future, more field survey will be done and more current data collected to consolidate the robustness of the model. The simulation covers a time span of 50 years and takes each year as time step. Next we will explain how to formulate the relationships among

factors and level and then articulate them logically and mathematically.

Population development sub model

The first sub model is the population development model. The factors that influence the level of population are shown in Figure 3. Population in the system is increased with the inflow of birth and decreased with the outflow of death and mortality. Fractional Birth Rate and Population determine the amount of Increase. The Fractional Birth Rate is further decided by the fraction of couples who are in the age of first time bearing child but never gave birth to any child before, the age at which the couple can be allowed to bear first child and the number of children that the couple can give birth to during the period of bearing age. Decrease presumably depends on the life expectancy in the country under investigation. The equations adopted in the sub model are listed as follows:

- (1) $\text{Population} = \text{Increase} - \text{Decrease}$
- (2) $\text{Increase} = \text{Fractional Birth Rate} \times \text{Population}$
- (3) $\text{Fractional Birth Rate} = \text{Fraction of First Child Bearing Age Couple} \times \text{First Child Bearing Age} \times \text{Number of Birth for One Couple}$
- (4) $\text{Decrease} = \text{Population} / \text{Average Lifetime}$

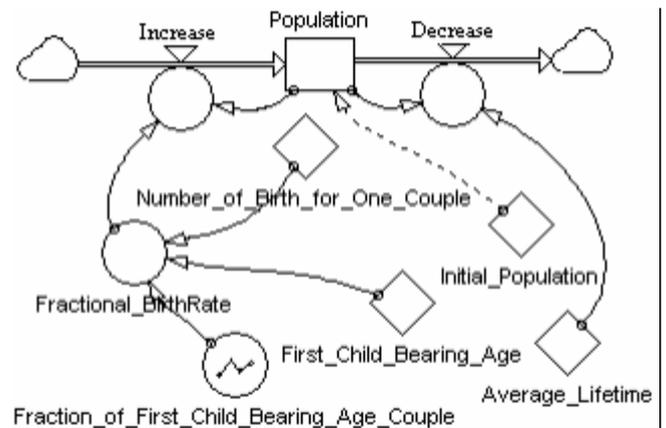


Figure 3: Population development sub model

Capable population conversion sub model

The second sub model is the model for determining the level of the capable population. The population is classified into two categories. One is the capable population which is an independent economic entity in the society and to whom mobile terminals are useful. Another part is the potential capable population who is a dependent economic entity and to whom a mobile terminal is temporarily useless. For example, young people under eighteen years old are typical potential capable population but not of the actual capable population. When they grow up they become members of the capable population. Elderly people over a certain age are presumably regarded as persons who will never use mobile terminals. Elderly people will be excluded from the initial value of the Capable Population. As depicted in Figure 4, the Historical Birth controls the addition to the Capable Population. After eighteen years,

the addition to the Capable Population is the today's Increase, an 18-year delay of Increase. To simplify our simulation, we choose eighteen year as time span for the first simulation. After the first simulation, we get the data about Increase. After we feedback the data to Historical Addition, we can get another 18 year simulation. As long as the simulation is needed, theoretically, the time span for Historical Addition can extend to any time period. Analyzing the historical birth number of the previous eighteen years and directly drawing a curve between birth and year can determine the addition rate. It is easily realized through lookup functions in the simulation, which allows manually inputting the collected data or directly importing data from other applications.

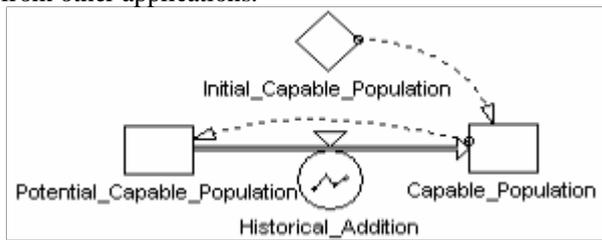


Figure 4: Capable population conversion sub model

Mobile terminal user addition sub model

The third sub model is the model for determining the level of mobile terminal users, i.e. modeling the addition of mobile terminal users in the mobile terminal market. We assume that personal income influences the addition of mobile terminal users. With the improvement of living standards brought by economy development, more and more people can afford to buy mobile terminals. Every year, a certain number of persons become actual mobile terminal users. In developing countries, there exists a great gap in the economic environment between urban and rural areas, so the addition is assumed to be different under rural and urban contexts. The concept of urbanization is introduced.

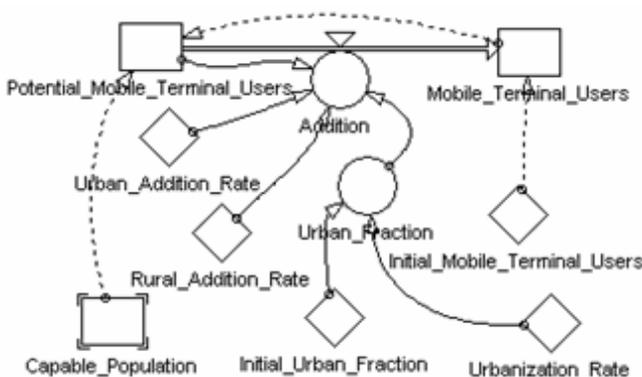


Figure 5: Mobile terminal addition sub model

As illustrated in Figure 5, Addition is determined by Rural Addition Rate, Urban Addition Rate and Urban Fraction. Urban Fraction is also a dependent variable

that depends on the urbanization pace in the region. The key equations in this level are expressed as follows:

- (1) Addition = (Rural Addition Rate x Rural Fraction + Urban Addition Rate x Urban Fraction) x Potential Mobile Terminal Users.
- (2) Rural Fraction = 1 - Urban Fraction
- (3) Urban Fraction = Initial Urban Fraction x (1 + Urbanization Rate).

Mobile commerce user conversion sub model

The final sub model is the model for determining the level of mobile commerce users, i.e. modeling the conversion of mobile commerce users or the diffusion of mobile commerce service in the mobile commerce market. Because mobile commerce service is a technological innovation, Bass diffusion theory [14] is applicable here.

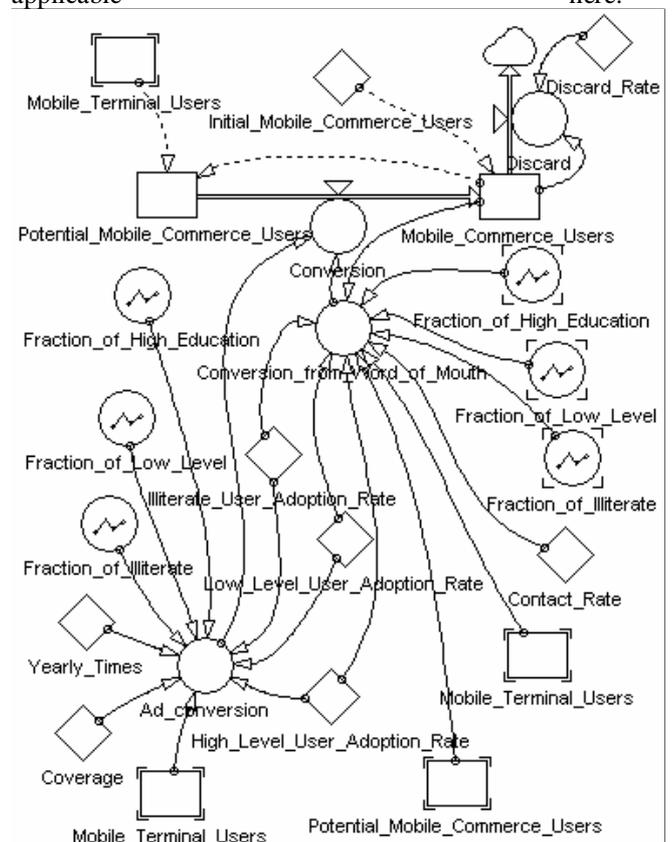


Figure 6: Mobile commerce user conversion sub model

The sub model in Figure 6 illustrates the factors that directly affect the conversion. The influences mainly come from the scale and intensity of information disseminations about mobile commerce among target users and the education level of recipients. The information propagation refers to advertising and "word of mouth". Based on the education level, mobile terminal users are classified in three groups. One is called illiterate who have received no education at all. Another one is defined as low level education users who

have finished the primary or secondary or equivalent education, and the other one is called high level education users who have achieved tertiary or tertiary-above education. Because they have different education background, they take different attitudes toward mobile commerce services. Their attitudes are reflected on their adoption rate.

The equations used in the model are listed as follows:

(1) Conversion = Ad conversion + Conversion from Word of Mouth - Discard

(2) Ad conversion = (Fraction of High Education x High Level User Adoption Rate + Fraction of Low Level x Low Level User Adoption Rate + Fraction of Illiterate x Illiterate User Adoption Rate) x Mobile Terminal Users x Coverage x Yearly Times

(3) Conversion from Word of Mouth = (Fraction of High Education x High Level User Adoption Rate + Fraction of Low Level x Low Level User Adoption Rate + Fraction of Illiterate x Illiterate User Adoption Rate) x Potential Mobile Commerce Users x Mobile Commerce Users/Mobile Terminal Users

(4) Discard = Mobile Commerce Users x Discard Rate

On the aspect of advertisement, the times and coverage area is controlled by mobile commerce providers. Every advertisement covers a certain number of populations, which consists of persons with three different education backgrounds. Advertising will induce a number of potential mobile commerce users to adopt the new service and become an actual mobile commerce user. Each kind of person with specified education background has its own fraction in target population and conversion rate. The total effect from advertisement is calculated using equation (2).

Another aspect is direct contact, also called demonstration effect or "band wagon". The process is visualized as someone in the Potential Mobile Commerce Users group being converted into an Actual Mobile Commerce Users. The two groups of people communicate and make contact with one another. When they make contact, there is some chance that the comments of the person who is a member of the Actual Mobile Commerce Users group will cause the person who is in the Potential Mobile Commerce Users group to adopt the service. The model shown in Figure 6 assumes that for each such contact between a person in the Actual Mobile Commerce Users and a person in the Potential Mobile Commerce Users there will be a number of conversions. This conversion is called conversion from word of mouth. As illustrated in equation (3), the number of conversion in a year will be equal to the sum of contributions from each category of potential mobile commerce users with different educational background. The conversion of each category user is equal to the specified adoption rate times the specified fraction in the Potential Mobile Commerce Users groups, times the product of Actual

Mobile Commerce Users and Potential Mobile Commerce Use, and times the Contact Rate.

Another important outflow in this level is Discard, which is shown in Figure 6 as a sink. Because mobile commerce is still in its infancy, there inevitably exist a lot of issues such as ease of use, security, usability, technological obsolescence etc, which bring discard and reduce the level of actual mobile commerce users. As expressed in equation (4), Discard is the product of discard rate and actual mobile commerce users. In reality, the discard rate may be very high at the beginning of conversion. As time elapses, the discard rate will decrease.

Mobile commerce diffusion model

The main model can be obtained by articulating the relationships among levels. Population equals potential capable population plus actual capable population. The actual capable population is the sum of potential mobile terminal users and actual mobile terminal users. Actual mobile terminal user involves potential mobile commerce users and actual mobile commerce users. Each level in the upper stream provides information to the next level.

3. SIMULATION AND FINDINGS

We collected preliminary data for the Chinese market available from the Chinese Statistics Bureau and the United Nations database. Wherever data was not available, we used common sense to arrive at some reasonable values. The results of a simulation run of a typical scenario with a time span of 50 years are shown in Figure 7 and compared in Figure 8.

The first graph in Figure 7 shows the population level, which grows very fast in the first 10 years. The level then reaches a peak of 1.5 billions after 10 years. Afterwards it starts to decline and eventually drops down to the initial level of 1.3 billions. The second graph shows the capable population level, which grows linearly from the initial amount of 600 millions to 1 billion in 50 years time. The third graph shows the mobile terminal users level which increases from the initial amount of 200 millions to 570 millions. During this period the market of mobile terminal users displays a goal-seeking behavior to reach a goal (arbitrarily set at 570 millions in the simulation run). The fourth graph at the bottom shows the level of conversion of potential mobile commerce users to actual mobile commerce users. This follows a quasi S-shaped pattern. In the first 30 years, the number of actual mobile commerce users will reach a maximum of 200 millions. The whole process will experience three stages. The first stage (market introduction) starts at the beginning of the simulation period and ends where the curve starts to take off. During this phase, which lasts for 15 years, development is slow as the number of actual mobile

commerce users is less than 30 millions. The second stage (market growth) covers the period from the time the exponential development starts to the time it starts to level off. This phase lasts for another 15 years. In this phase, the number of actual mobile commerce users skyrockets from an initial amount of 30 millions to 200 millions at the end. The increase is initially exponential and then changes to a goal-seeking pattern. In the final stage (market maturity), the dynamic market, after booming in stage two, reaches a new equilibrium. The amount of actual mobile commerce users is stabilized and remains at a high level.

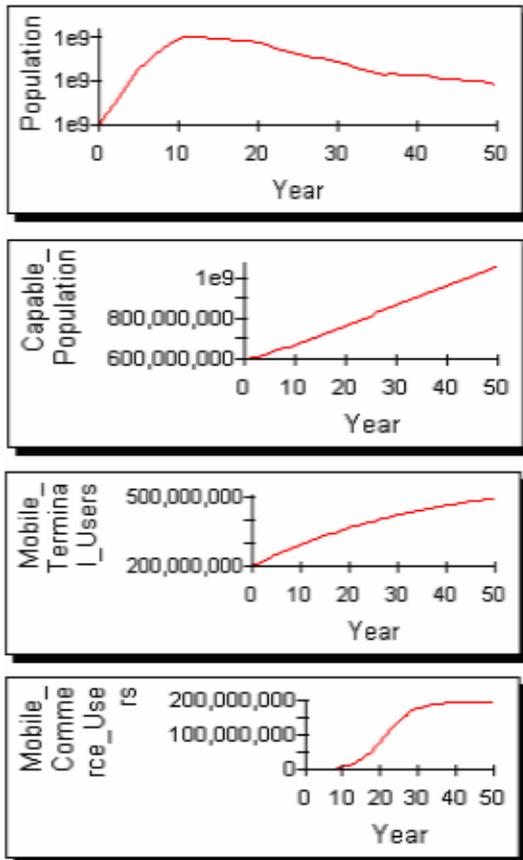


Figure 7: Key levels change in the system

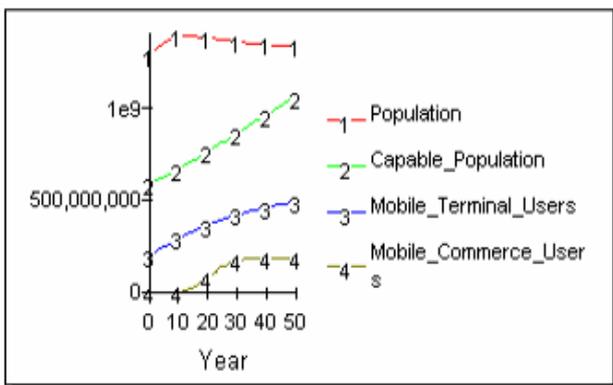


Figure 8: Contrast of the key level development in China

4. CONCLUSION

We have proposed a system dynamics model of mobile commerce diffusion and used preliminary data from the Chinese market to perform some simulation runs. Our intention is not to forecast the number of mobile commerce users in the future. Rather, our purpose is to discover the factors that influence the development of mobile commerce and their inter-relationships. This will undoubtedly help in decision-making processes such as exploring the effects of different policies and strategies to achieve certain goals (e.g. a targeted number of users, etc) under different scenarios. The model is still under development and we need to collect more data from the Chinese market (telephone interview, questionnaire survey, etc.) to improve it. We are also considering conducting a Delphi survey for the purpose of validating the model. Experts will be recruited from China and their feedback on the model sought.

REFERENCES

- [1] E. F. Wolstenholme, *System Enquiry*: John Wiley & Sons, Chichester, 1990.
- [2] M. C. Jackson, *Systems Thinking: Creative Holism for Managers*. University of Hull, UK: John Wiley & Sons, Ltd, 2003.
- [3] J. W. Forrester, "Counterintuitive Behavior of Social Systems," *Technology Review*, 1995.
- [4] J. D. Sterman, *Business Dynamics: Systems thinkings and modeling for a complex world*. Massachusetts: The McGraw-Hill Companies, Inc., 2000.
- [5] E. F. Wolstenholme, *System Enquiry: A System Dynamics Approach*. Chichester: John Wiley & Sons, 1994.
- [6] R. G. Coyle, *System dynamic modelling: A practical approach*. London: Chapman & Hall, 1996.
- [7] C. W. Kirkwood, *System Dynamics Methods: A Quick Introduction*. College of Business, Arizona State University, 1998.
- [8] J. W. Forrester, *Industrial Dynamics*: MIT Press, Cambridge MA, 1961.
- [9] J. W. Forrester, *World Dynamics*. Cambridge, Massachusetts: Wright-Allen Press, Inc., 1973.
- [10] J. W. Forrester, *Urban Dynamics*. Cambridge: The MI. Press, Massachusetts Institute of Technology, 1969.
- [11] M. L. Deaton and J. J. Winebrake, *Dynamic Modeling of Environmental Systems*. New York, 2000.
- [12] E. M. Rogers, *Diffusion of innovations*, 3 ed. New York: The Free Press, 1983.
- [13] M. A. Quaddus, *Diffusion of information technology: a review of static and dynamic models (working paper)*. Perth, Western Australia: Curtin University of Technology, 1995.
- [14] F. Bass, "A new product growth model for consumer durables," *Management Science*, vol. 15, pp. 215-227, 1969.

From Control to Competition: Trust Institution in E-Business

Yunfu Huo, Deli Yang, Jianjun Wang, Qiong Guo

School of Management, Dalian University of Technology, Dalian, Liaoning 116024, China
somdyang@dltu.edu.cn

ABSTRACT

Global economy has formed a huge challenge to the trust of traditional business. It's a important problem of all enterprise how to build a effective trust institution in the new economy. The rapid development of E-business is bringing a turning point of trust institution. In traditional trust institution, trust means that principal controls the agent in fact. The evolution of trust institution was closely related to the development of technology because of control. This paper presents, with the innovation of technology and system, precondition of trust through the method of controlling has already been out of the existence. The development of E-business will have influence on the evolution of trust institution, Advantages of E-business can promote monitor from man-made to market-made, from control to competition. It will build an effectual trust institution.

Keywords: trust, control, competition, E-Business, advertising game

1. INTRODUCTION

Global economy has formed a huge challenge to the trust of traditional business. It's an important problem of all enterprise how to build an effective trust institution in the new economy. The rapid development of E-business is bringing a turning point of trust institution.

2. THE PROBLEM AND CHALLENGE

2.1 The traditional trust

Trust is a definite expectation of other's action (Gambetta, 1988) as well as an active ability that engage oneself to accept the many social institution (Zucker, 1986). Trust is necessary to reduce transaction cost, and help business organizations operate better (Miles & Snow, 1992). Trust becomes a pivotal factor of the thought and philosophy of management (Miles & Greed, 1995). Trust is rooting in three patterns: process based trust, character based trust, and institution based trust (Zucker, 1986). Institution, especial comparability and reciprocal experience decide the trust (Arrow, 1984). Trust was a situation that we relied on others; it meant dependence together with risk. Because calculativeness is ubiquitous (Williamson, 1998), the trust of coming from dependence means that principal controls the agent in fact. Trust is derived from distrust. The premise of trust is the exchange of direct interaction and sensing knowledge. So some scholars think that management is one art of control (Putti & Weihrich & Koontz, 1998).

2.2 Trust and technology

The evolution of trust institution is closely related to the development of technology because of control (Kipnis, 1991). The development of technology is divided into three stages: skills based technology, which corresponding to the era of producing by hand; the routine technology, which corresponding to the era of

producing by machines; informational technology, which corresponding to the era of producing by intelligence (Faunce, 1981). The principal cannot only simply rely on the agent's moral because of the complicated technology. The dependence is more displayed that whether the agent can better deal with the technology, at this moment, the need of agent's skills and moral improved highly. It was more difficult to control than before. Further more the way that technology used make the implied terms of traditional trust out of existing, more and more trades based on multi-region and provisional colony were substituted trades of self-region and steady colony (Kipnis, 1994), codified knowledge which is gradually replacing the sensual knowledge is becoming basis of the trade (Masahiko Aoki, 2001).

3. TRUST IN E-BUSINESS: AN ADVERTISING GAME MODEL

Because the traditional trust institution has relied on strong ethical hypothesis, it can be described as a game model. Players have all or part knowledge of the game structure. The trust institution is a common strategy choice of players, and it can be self-enforcing. We will analyze the trust institution by an advertising game model that is reformed by Spence's job market model (Spence, 1974).

3.1 Advertising game model

Because of asymmetric information, bargainors know the quality level of their goods, buyers don't. But the information of quality can be transferred to buyers by advertising. According to Spence-Mirrlees condition (Spence & Mirrlees, 1974), there are different costs of advertising between different quality. The cost of advertising is inverse ratio of quality. The efficiency improvement of different quality is different.

Thinking of a buyer and a bargainer. Quality has two estates: if $\theta = 1$, the quality is low, if $\theta = 2$, the quality is high. The bargainer is clear of the real value of θ , but the knowledge of the buyer is only the probabilities of quality. Let the probability is p : $p(\theta = 1) = p(\theta = 2) = 1/2$. Before contracting, the bargainer will select advertising level. Let the level is s : $s \in \{0, 1\}$, $s = 0$ means advertising, $s = 1$ means no advertising. The cost of advertising is $C(s, \theta) = s/\theta$. After advertising is observed, the buyer will decide the price $w(s)$. The bargainer will choose whether receiving the price or not. If being received, the buyer has an expected value: $y = \theta$ (it is a hypothesis that advertising isn't influence of value), the bargainer's expected utility is $U(s, \theta) = w - s/\theta$. The buyer's expected utility is $\pi(s, \theta) = \theta - w(s)$. Otherwise, $U = \pi = 0$. At one time, it is a hypothesis that there is a complete market, so when the market has an equilibrium, price will be equal to expected value and the buyer's expected utility is 0.

3.2 The game of trust in control

The analysis of this game model has two angle of view: symmetric and asymmetric information games. By symmetric information, advertising needs many costs but has no value. So whether high with low of the quality, bargainer will choose $s = 0$, the price of low quality is $w(\theta = 1) = 1$, the price of high quality is $w(\theta = 2) = 2$. There is trust institution in the market this time. But this Pareto dominance equilibrium will not exist in asymmetric information games. If the buyer doesn't know θ , his expected value will be $y = 0.5 \times 1 + 0.5 \times 2 = 1.5$, the competition of players will have a result in $w = 1.5$. But, when the advertising can transfer quality signal to buyers, $w = 1.5$ would not able to be Nash equilibrium (Zhang, 1996).

Traditional trust institution is built by a hypothesis of symmetric information games. Besides of a strong common share of the institution, it also needs a small self-region market. Along with the development of technology, the bound of market is more and more expanding. The hypothesis of traditional trust institution is removing from the fact. For obtaining the symmetric information, many other forcing controls are necessary, such as quality standard, understanding of craftworks and abilities, etc. In fact, the trust has been equal to the control. The essence of "dependence and control" has brought high cost, and can't meet the demand of trust in the times of global economy. The principal cannot far directly control the behavior of the agents because of his bounded rationality. Controlling has sharpened the contradiction between the principal and agent.

3.3 The game of trust in competition

Though the traditional business environment has stronger regional characteristics, the conditions of the hypothesis of the trust are easier to realize, the controlling, as a kind of ways to form trust, can work in a certain extent. With the innovation of technology (modern manufacturing, modern transportation and modern information) and system (the freedom of trade), the modern economy becomes globally day by day, precondition of trust through the method of controlling has already been out of the existence. Under this environment, the agent who is prone to calculativeness will not promise openly, so the effective trade can only be realized on terms that the trust is restrained from reliably supporting (Williamson, 1998). As the important displaying form of the technology at the third stage, E-business is not only the handed technology, but also is the main means to change the environment. The development of E-business will have influence on the evolution of trust institution.

In E-Business, by the regional difference of culture and moral, the belief of traditional trust institution will not be more and more common shared. Because of the asymmetric information of quality, buyers can only observe s not θ , so the price will only be decided by s . Let $\mu(\theta = 1 | s)$ is the buyer's posterior probability of low quality when he has observed s that the bargainer selected. Perfect Bayesian equilibrium means: (1) the bargainer selected $s(\theta)$, (2) the buyer knows $\mu(\theta = 1 | s)$ and decides $w(s)$ by the "s" that he has observed. The result is: ① by $w(s)$, $s(\theta)$ is the best choice of the bargainer when quality is θ ; ② by $s(\theta)$, $\mu(\theta = 1 | s)$ is accordant to Bayesian ruler, $w(s)$ is the best choice of the buyer.

The equilibrium may be pooling or separating. First, pooling equilibrium (PE) means the bargainer selected same advertising in different qualities and won same price. By $s(\theta) \equiv 0$,

$$(PE): \begin{cases} s(\theta = 1) = s(\theta = 2) = 0 \\ w(0) = w(1) = 1.5 \\ \mu(\theta = 1 | s = 0) = 0.5 \\ \mu(\theta = 1 | s = 1) = 0.5 \end{cases} \quad (1)$$

It means no advertising of any qualities is Pareto dominance equilibrium. The buyer considers there is no quality signal in advertising, so the price is equal to the expected utility. Because of the buyer's Pareto dominance equilibrium is $w \equiv 1.5$, $C(1, \theta) > 0$, so $s(\theta) = 1$ is impossible, the bargainer's Pareto dominance equilibrium is no advertising ($s(\theta) = 0$) (Zhang, 1996).

Because we assume the buyer will not rework his prior probability when he finds $s(\theta) = 1$, (PE) is Pareto

dominance equilibrium. The invariable prior probability in advertising game demands players have share-beliefs of sensual knowledge and codified knowledge, by all appearances, it could not come to existence in global economy and E-business. If the buyer's posterior probability is $\mu(\theta = 1 | s = 1) = 0$, it means the quality advertised must be high, the (PE) will not come into existence. For by $\mu(\theta = 1 | s = 1) = 0$, when the bargainer's choice is $s(\theta) = 1$, the buyer's choice will be $w(1) = 2$. The high quality will select advertising, $U(s = 1, \theta = 2) = 2 - 1 / 2 = 1.5$. No advertising is poor, $U(s = 0, \theta = 2) = 1 - 0 = 1$, so we will have the separating equilibrium (SE):

$$(SE) : \begin{cases} s(\theta = 1) = 0, s(\theta = 2) = 1 \\ w(0) = 1, w(1) = 2 \\ \mu(\theta = 1 | s = 0) = 1 \\ \mu(\theta = 1 | s = 1) = 0 \end{cases} \quad (2)$$

(SE) is perfect Bayesian equilibrium. When the buyer has selected the posterior probability and the price, if $U(s = 1, \theta = 2) = 1.5 > U(s = 0, \theta = 2) = 1$, the quality is high, the best strategy is advertising; if $U(s = 0, \theta = 1) = 1 \geq U(s = 1, \theta = 1) = 1$, the best strategy is no advertising when the quality is low. Otherwise, by the bargainer's choice, the buyer's posterior probability is based on Bayesian ruler, so the price strategy is Pareto dominance, and other (SE) isn't being (Zhang, 1996).

4. CONCLUSIONS

The advertising level is a signal in (SE) which transfers information of quality to buyers. Spence-Mirrlees condition is the precondition of this game equilibrium. When the same advertising is selected, the cost of high quality is lesser than low quality, the high quality will have enough incentive to separate itself from others by advertising. Otherwise, for winning same price, the low quality will have enough incentive to simulate the high quality by same advertising strategy. Thus the buyer will not believe the advertising as a quality signal. It will bring on market failure if there is not enough quality information in other ways. The trust isn't being. So we will see the conclusions as follows:

1. In global economy and E-business, because of scarcity of sensual knowledge, the condition of strong ethical hypothesis in traditional trust institution is not being. The new trust's game model should be more and more based on "an invisible hand" (Smith, 1776). The ethical hypothesis is feeble.

2. In E-Business, Spence-Mirrlees condition makes it clear that the system assuring of real advertising is necessary. If the discount factor $\delta > 0$, players will have enough incentive to tell the truth in infinitely repeated games (Rasmusen, 1994). It is necessary to

build and improve the stable common E-Business market by strong legal system. The third party will regulate players' actions in market, such as quality, reality of advertising, etc. The certification authorities and the continuous open newsreel of business can make the cost is too expensive to lie.

3. If the bargainers' quality information can help buyers purchase properly, advertising can improve the effect of resource allocation. By taking the advantages of E-business that can help to share the mass codified knowledge, purchase and supply on-line can bring an effect of vendue. The good-right order can translate from traditional mode such as prior plan, unite control, posterior mediation, into the mode of contract that based on real product (Masahiko Aoki, 2001). Lesser asymmetry and uncertainty information will make enough incentive of the bargainer to improve quality. At one time, opening competition can promote monitor from man-made to market-made by the endogenous punishment mechanism by constantly repeated games. It is from control to competition.

ACKNOWLEDGEMENT

This paper is supported by the National Natural Science Foundation of China (No: 70031020).

REFERENCES

- [1] Arrow, K., *The limits of organization*, New York: Norton, 1984.
- [2] Faunce, W.A., *Problems of an industrial society*, New York: McGraw-Hill, 1981.
- [3] Gambetta, D., Can we trust trust? In D. Gambetta (Eds.), *Trust: Making and breaking cooperative relations*. Oxford, UK: Basil Blackwell., 1988.
- [4] Kipnis, D., "The technological perspective", *Psychological Science*, No. 2, 1991.
- [5] Kipnis, D., "Accounting for the use of behavior technologies in social psychology", *American Psychologist*, 49, 1994.
- [6] Masahiko Aoki., *Towards a Comparative Institutional Analysis*, Shanghai: Shanghai Far East Publishers, 2001
- [7] Miles, R.E., & Snow, C.C., "Causes of failure in network organizations", *California Management Review*, Vol. 34, No. 4, 1992.
- [8] Miles, R.E., & Greed, W.E.D., "Organizational forms and managerial philosophies: A descriptive and analytical review", *Research in organizational behavior*, Vol. 17, 1995.
- [9] Putti, J.M. & Wehrich, H. & Koontz, H., *Essentials of Management: An Asian Perspective*, Singapore: McGraw-Hill. Book Co-Singapore., 1998.
- [10] Rasmusen, Eric, *Games and Information: An Introduction to Game Theory*, Cambridge: Blackwell Publishers, 1994.
- [11] Smith, A., *The Wealth of Nations*, Shijiazhuang: Hebei Science & Technology Publishers, 2001.

-
- [12] Spence, A. M, "Job Market Signaling", *Quarterly Journal of Economics*, Vol .87, 1974.
- [13] Williamson,O.E, "Calculativeness, trust, and economic organization", *Journal of Law and Economics*, Vol .34, 1998.
- [14] Zang, Weiyang, *Games and Information Economics*, Shanghai : Shanghai people's Publishers, 1996.
- [15] Zucker,L.G, "Production of trust: Institutional sources of economic structure, 1840~1920", *Research in organizational behavior* , Vol .8, 1986.

Negative Reputation Rate as the Signal of Risk in Online Consumer-to-Consumer Transactions

Dahui Li¹, Zhangxi Lin^{2*}

¹ Labovitz School of Business and Economics, University of Minnesota Duluth, Duluth, MN 55812, USA,

² College of Business Administration, Texas Tech University, Lubbock, TX 79409, USA

¹dli@d.umn.edu, ²zlin@ba.ttu.edu

ABSTRACT

Previous online reputation research has been focused on the effects of positive and negative reputations on trust formation, trading price, and probability of sale. We propose that negative feedback rate (NFR) is the most important indicator of the risk of buying from a seller online. This proposition has been supported by an empirical study based on data collected from eBay.com. We found that the 6-month NFR in the current period predicts much better the future risk measured by the NFR in the next 6-month period than did net reputation score and negative reputation score. A seller's life-long negative score in fact was not significant in predicting the future risk. In addition, a seller's age in the market was found to have similar predicting power on risk as did net reputation score.

Keywords: Reputation; Electronic commerce; Risk; C2C auction

1. INTRODUCTION

The prevailing online reputation systems, which provide trust-building mechanisms and assurance services on eBay, Yahoo auction, Amazon and many other online auction marketplaces, have contributed significantly to the success of the consumer-to-consumer (C2C) electronic commerce. The online reputation systems allow traders to leave each other a positive, negative or neutral rating after a transaction is finished. The scores are then added up as quantitative indicators of traders' reputation records which are available to future traders. In addition, traders can leave brief comments on their transactions or trading partners for further references.

The effectiveness of online reputation systems in the C2C auction market has triggered wide interests from the research community. According to the synthesis by Dellarocas [4], although some have investigated the effect on the probability of selling products [2], empirical studies on online reputation are unanimously focused on the effect of seller's feedbacks on item prices. However, because of the diverse nature of research design, such as different types and categories of items being investigated, different measures of reputation, different data collection processes, and different sample sizes, no consistent findings were found across these studies [13]. Positive feedback might or might not increase price or probability of sale, and the same applies to negative feedback and net reputation score, as Resnick et al. [13] summarize from 16 research reports. Whereas, accumulated positive and negative scores, as well as negative scores from a period of time, were found to be the most influencing

components of seller's reputation on buyer behavior.

On the other hand, some research has investigated the effect of risk or perceived risk on online trading and the adoption of trusted third party services (e.g., Antony et al. [1]; Hu et al. 2001 [7]). Perceived risk was identified as an important determinant of purchasing items and adopting trusted third parties' services in the electronic market. Further, perceived risk is subject to change with regard to each trading partner's online reputation.

The basic reputation indicators that can be obtained from a C2C online reputation system include positive score, negative score, and neutral score (less often used). Based on these three basic scores, three combination scores can be derived: the *total reputation score* which is the sum of positive, negative and neutral scores, the *net reputation score* which is the difference of positive score and negative score accumulated from the ratings from unique traders, and the *negative feedback rate* which is the ratio of negative score and the total reputation score. Presently, a majority of reported research uses positive, negative, and net reputation scores as the indicators of online reputation, but NFR has not been touched yet.

How do these reputation scores reflect the risk of trading with a seller? To our knowledge, little research has answered this question. So far, majority research effort is in the effect of reputation on trust, but not aimed at the measure of the risk in the C2C online transactions. The main research idea in this paper is described in Figure 1. In the dotted box of the figure is a typical model used in trust related research (see [8]),

* Also an adjunct professor, Fujian University of Technology, Fuzhou, China

where perceived risk is investigated. However, the perceived risk does not necessarily match the real risk in the future, because online traders' behavior could be irrational. Therefore, this paper is intended to investigate the relationship between reputation indicators and the risk. We propose that a seller's negative feedback rate (NFR) is the most important indicator for predicting the risk of buying from him. Based on data collected from eBay, we conducted an empirical study. The result supported our proposition.

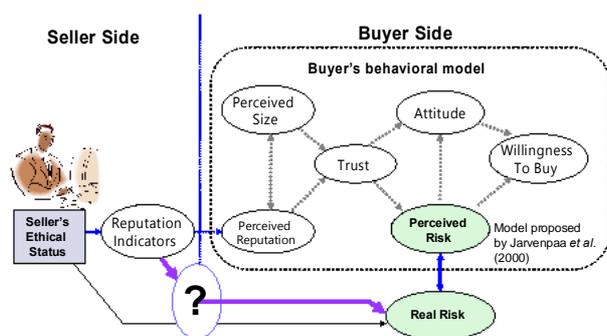


Figure 1: A conceptual model of the research

2. THEORETICAL BACKGROUND

2.1 Reputation

In the traditional reputation literature, reputation is generally defined as the consistency of an entity's behavior over a certain period of time [3][5][14]. It is a record of the history of an entity's interactions with others. Thus, reputation building is not a one-time effort but is based on the sum of all the past behaviors of the entity.

Reputation can be positive or negative. A positive reputation manifests all the favorable assessment of an entity, while a negative reputation shows the unfavorable aspects of the entity. Because the potential sacrifice associated with a negative or bad reputation is very high, an entity with a positive reputation is predicted to behave consistently in a favorable manner in the future. In buyer-seller relationships, the seller's reputation has a positive effect on buyer's trust in the seller and buyer's long-term orientation with the seller [6].

In electronic commerce research, Internet buyers are found to favor web sites that sell familiar products manufactured by familiar merchants [12]. The reputation of an online store is positively associated with an online consumer's trust in the store [8]. Further, research has shown that the negative reputation has much more effect than the positive reputation on buyers' trust because the negative reputation is the repeatedly appearance of a seller's unfavorable behavior that is highly associated with the potential trading risk in the future (See the review from Dellarocas [4] and Resnick et al. [13]).

2.2 Risk

According to classical decision theory, risk is "the variance of the probability distribution of possible gains and losses associated with a particular alternative" ([11], pp.1404). The definition is too abstract for the traders in the C2C auction market to understand. Following MacCrimmon and Wehrung [10], risk is defined as the chance or probability of loss. We thus define the potential risk of trading with a seller to be the probability of leaving the seller a negative feedback. Although the reputation system provides detailed comments about the negative reputation, a potential buyer may not have the resources (cognitive, motivational, or other external resources) to search the entire archive and find out the exact reasons for the negative comments. Further, risk is also associated with the significance of loss [11], e.g., how much a buyer can lose because of fraud.

2.3 The rationale of the research

In this research we focus on how online negative reputation can indicate the real risk in a future online transaction. This issue is represented by the question mark in Figure 1. Normally, a reputation system does not provide a direct causal link between the historical negative feedbacks and the information about the auctioned item, so that a buyer could not easily figure out the probability of future loss. Hence, we argue that a potential buyer will rely on the summarized reputation information in a seller's profile to judge the risk level of trading with the seller. The definition we have provided implies that risk is in the form of probability rather than a quantitative number representing the value of the loss. In the C2C online auction context, previous research has used negative feedback scores to measure risk. This measure is not accurate, because a negative feedback score does not reflect the probability of getting negative feedback in the future. The right measure of such risk should be a number signaling the probability of unpleasant consequences from the transaction with the involved seller. Thus, both the number of negative feedbacks the seller has received in a period of time and the total number of transactions in the same period of time should be considered together. We realize that there are several different types of NFRs, e.g., lifetime NFR and NFRs during different time periods. They are applicable to the same risk analysis purpose depending on different timeframes and accuracies. In the later discussion, we refer NFR as the NFR calculated from reputation scores in a 6-month period from now on.

3. AN EMPIRICAL STUDY

In this section, we report an empirical study based on data collected from eBay.com. The purpose of the empirical study is twofold. First, we aim to reveal the nature of the negative feedbacks through a content analysis of the feedback comments. We have not seen

any effort of analyzing the diverse contents of the reputation feedbacks. Second, we want to justify the predictive power of different reputation scores on the risk in the transactions with unknown sellers through a longitudinal study, using reputation score data collected at two time points.

3.1 Study 1 - Content Analysis of Negative Feedback Comments

For the content analysis, we randomly collected 216 unique sellers with negative reputation scores. We searched these 216 seller’s transactional history archived on eBay.com. We randomly selected one piece of feedback comment associated with a negative rating to the seller.

As each online negative feedback may link to different stories and signal different types of perceived risks, such as financial losses, undesired low quality, and unsatisfied service, it is necessary to look into negative complaints in order to reveal their nature. We are interested in why buyers are unhappy with sellers and how buyers attribute those complaints.

The authors of this paper worked as independent coders of the feedback comments. After finishing coding two coders compared the results and resolved the differences and discrepancies with a detailed discussion. The reliability of inter-coders was 0.90.

The coding was focused on three basic questions: What are the causes of negative feedback, merchandize or service? Who are responsible for the negative feedback, the seller, the buyer, or the third party? Is the negative feedback resolvable?

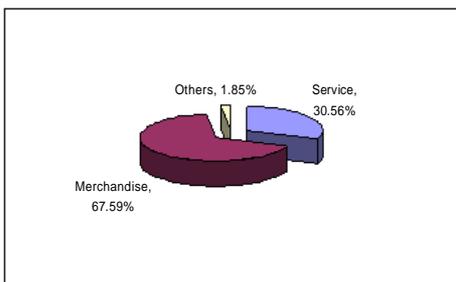


Figure 2: Causes of complaints

Results and Implications

The results of the coding on the three questions are shown in Figure 2-4 and Table 1. The three findings are discussed as follows:

Not every complaint is related to a merchandize problem

Figure 2 indicates that about two thirds of negative comments are complaints against merchandize problems. More than 30% complaints are related to seller services,

such as “shipped to wrong address”, “item received more than a month”, or “never response to my email”. Apparently these complaints are not related to any fraud, but reflect buyers’ unpleasant experience. Thus, this type of risk is about unsatisfactory transaction that every buyer wants to avoid.

According to the responses from the seller, the buyer could be the number one trouble maker followed by the seller and the third party.

There are many attribution errors in the negative feedbacks left by the buyer. According to Table 1, more than one third of complaints are because of the buyer. Typically, an inexperienced buyer may not understand the information of the auctioned item well and does not follow the way as an experienced buyer does. This is particularly remarkable in the complaints about the merchandize. Although this is not the kind of the risk in the traditional sense, the over reaction of a buyer may incur the retaliation of the annoyed seller returning with a negative rating. So the backfiring from the seller or uncomfortable feeling of the regrettable outcome is the risk the buyer may face under this situation.

Table 1. The liable agents to the complaints

Complained issues	Seller	Buyer	3rd party	Others
Merchandise	18.49%	38.36%	15.07%	28.08%
Service	12.12%	28.79%	19.70%	39.39%
Overall	16.20%	36.11%	16.20%	31.49%

In addition, about one sixth of the complaints are related to the third party who provides delivery services, contracted packing services, or original product supply (Figure 3). Although this also has nothing to do with fraud, it is one kind of risks that a buyer may encounter. As shown in Table 1, sellers are responsible for less than half of the complaints as buyers are.

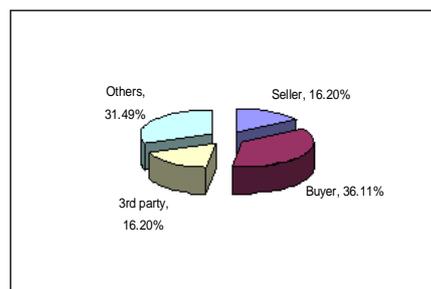


Figure 3: The composition of liable agents

More than 40% of the complaints are resolvable or have been resolved at the moment they are made.

Figure 4 shows that more than half of the complaints can be classified as disputes. Sellers are responsible for almost half of complaints (47.95%) to merchandises. Less than one fifth (18.49%) of the complaints against the merchandize has caused sellers’ harsh responses and

ended unpleasantly. Practically, when a seller has responded to a buyer’s complaint regarding the merchandise and a dispute is incurred, it becomes difficult to justify whether the case is a fraud or not. From a prospective buyer’s angle, this is an undesirable situation and must be avoided.

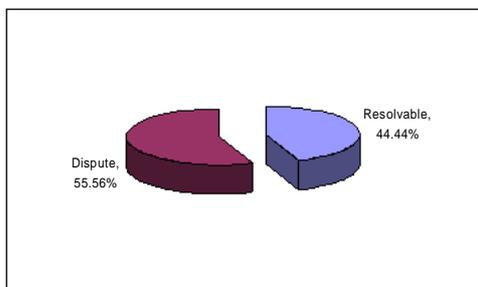


Figure 4: The resolvability of negative comments

3.2 Study 2 - Predictors of NFR

For the longitudinal analysis, we followed the data collection procedure reported in Lin et al. [9]. We first

collected 2,000 seller’s reputation data in March 2003. Six months later, we repeated the data collection using the same 2,000 seller IDs. 174 sellers IDs were no longer active in the second round collection because these sellers have quitted the account, leaving 1,826 useful observations in the study.

Results

We tested how well the reputation indicators observed in the current 6-month period (t0), i.e., NFR, net reputation score, accumulated negative feedback, and age, can predict an individual seller’s NFR in the next 6-month period (t1). We used stepwise regression to examine the correlation of the NFR in the next period to different reputation indicators. The results from the Tobit regression model and the OLS model are summarized in Table 2. Tobit regression model was used because the dependent variable NFR at the second point (t1) was censored with a lot of 0’s. In general, the regression results from the two models are consistent.

Table 2: Regression analysis of predictors of NFR in the next 6-month period
Dependent: NFR(t1)

Tobit Model	Ln(NFR(t0))	Age	Ln(Net(t0))	Ln(Neg(t0))	N	Log Likelihood
Model 1	0.87				1826	-640.40
Model 2		-0.0009			1826	-1281.76
Model 3			-0.40		1826	-1205.80
Model 4				-0.01ns	1826	-1031.06
Model 5	0.82	-0.0004			1826	-627.03
Model 6	0.78		-0.15		1826	-627.55
Model 7	0.89			-0.11 ns	1826	-633.27
Model 8	0.78	-0.0003	-0.10		1826	-622.96
Model 9	0.84	-0.0004		-0.06 ns	1826	-625.38
Model 10	0.69	-0.0003	-0.22	0.12 ns	1826	-621.50
OLS	Ln(NFR(t0))	Age	Ln(Net(t0))	Ln(Neg(t0))	N	R ²
Model 1	0.87				468	0.5854
Model 2		-0.001			468	0.1252
Model 3			-0.44		468	0.2408
Model 4				0.01 ns	468	0.0002
Model 5	0.82	-0.0004			468	0.6084
Model 6	0.78		-0.15		468	0.6076
Model 7	0.89			-0.11 ns	468	0.5979
Model 8	0.78	-0.0003	-0.10		468	0.6152
Model 9	0.84	-0.0004		-0.06 ns	468	0.6112
Model 10	0.69	-0.0003	-0.22	0.12 ns	468	0.6176

ns: non-significant; All the other coefficients are significant at the 0.001 level

Implications

Seller’s current NFR is the most significant variable for predicting the NFR in the next period

The most important predictor of future risk is the NFR in the current 6-month period, which explains the highest percentage of the variance of the next period

NFR (R² is about 59% from the OLS regression) among the four independent variables. This means that the NFR is the best among all four variable in predicting the risk of trading with an unknown seller in the C2C online auction market. The positive relationship suggests that the higher the NFR from the current period, the higher the NFR in the next period. The contribution of the other three variables in explaining the risk level is

marginal because their presence has little effect on the goodness of fit (for example, 63% vs. 59% of R^2 in the OLS regression, or -640 vs. -627 of logarithmic likelihood in the Tobit regression).

Net reputation score is negatively correlated to seller's NFR in the next period

The relationship between net reputation score and NFR is significantly negative, suggesting that the higher the net reputation score, the lower the NFR in the next period. This finding is consistent with the findings reported in previous literature, in which net reputation score has a significant effect on trust and price premium. Therefore, our study suggests that net reputation is also a significant predictor for the trading risk with a specific seller. However, the explanatory power of net reputation score (R^2 is about 24% in the OLS regression) is not as strong as NFR.

An explanation to the discrepancy of the effect of net reputation score is that when the net reputation score is high, the variance of NFR is low, because sellers with high net reputation scores tend to have high transaction volumes and their NFRs are relatively stable. The sellers with low net reputation score typically have lower transaction volume, and the variance of their NFRs is higher. Therefore, the NFR is more sensitive to the range of net reputation scores.

Seller's age matters in predicting NFR

The age of a seller also provides the similar explanatory power in predicting NFR (R^2 is about 13% in the OLS regression) as the net reputation score. This implies that the longer a seller stays on the market, the lower the NFR. Therefore, like net reputation score, a seller's age can be another important measure of his experience and tenure in the market. According to Lin et al. [9], in the context of market structure research, online reputation as the tangible asset for online traders reflect their capacity as virtual firms in the electronic market. Based on this, our study is the first one that has found that age is an effective measure of a seller's capacity in doing business on the market.

A seller's total negative score accumulated in his business life cycle has no effect on NFR

Unlike previous studies who have found significant effects of negative feedback on person's trust and price, this study did not find significant effect of negative reputation score on the next period NFR ($p > 0.05$). This suggests that without considering the total number of transactions from which the negative reputation score are generated, the risk in a trading cannot be predicted accurately. Negative reputation score, an absolute number rather than a relative number, cannot tell exactly whether it is likely to have an unpleasant outcome in the transaction with a seller. The previously reported significance of negative reputation score on traders'

behavior was only the implication of their irrationality.

According to the above findings, the maturity of a seller, signaled by his age and net reputation score, is important for estimating the risk level in transacting with him. We can refer to this as the "learning effect." The learning effect was also found from the negative relationship between transaction volume and negative rate ($\beta = -0.0002$, $p < 0.001$, $R^2 = 0.10$). The higher the number of transactions, the lower the NFR. Although an individual may maximize his effort to conduct a good business, his behavior is relatively consistent. If he has got higher NFR in the past, he is more likely to get higher NFR in the future. This implies that even though a trader changes his ID or identities on eBay, if he has got a high NFR, he is still likely to get the same level of NFR under a new ID, because his ability of doing business behind the pseudonym is unimproved. At the same time, the costs of switching ID and building new reputation may be high, which suggests to the sellers: *Do not change your ID when you have a high NFR. It does not work.*

4. CONCLUSIONS

In this paper, we analyzed the negative feedbacks by coding the textual complaint data and examined the determinants of future NFR with the historical reputation scores. We made two clear contributions to the research on C2C auction market. First, we revealed the diverse nature of negative reputation. Several important findings were derived from analyzing the content of the negative feedback comments. Second, we proposed to use NFR to measure the potential risk in trading with an unknown seller in the electronic market. The results of the empirical study could provide guidelines for the future buyers to do business online. The findings also have important implications for the design of reputation systems. Online companies are encouraged to report NFR as a measure of reputation of traders. We have noticed that since 2004 eBay.com has started to provide positive feedback rate on its online reputation forum (see Appendix). We reason that the positive feedback rate (PFR) might work as an incentive for sellers to maintain their reputation and promote the effectiveness of eBay's feedback forum. However, PFR is not as straightforward as NFR as an indicator of the risk, although rational buyers can derive NFR from PFR. If eBay.com can further show the positive feedback rate during a period of one month and 6-month, the figures will be more informative and valuable.

REFERENCES

- [1] Antony, S., Z. Lin, and B. Xu, "Determinants of Online Escrow Service Adoption: An Experimental Study," *The Proceedings of the 11th Workshop on Information Technology and Systems (WITS'01)*, 2001, pp71-76.

[2] Ba, S., and P. A. Pavlou, "Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior," *MIS Quarterly* 26(3), 2002, pp243-268.

[3] Balachander, S., "Warranty Signaling and Reputation," *Management Science* 47(9), September 2001, pp1282-1289.

[4] Dellarocas, C., "Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science* 49(10), 2003, pp1407-1424.

[5] Fombrun, C., and Shanley, M. "What's in a name? Reputation Building and Corporate Strategy," *Academy of Management Journal* 33(2), June 1990, pp233-258.

[6] Ganesan, S. "Determinants of Long-Term Orientation in Buyer-Seller Relationships," *Journal of Marketing* (58:2), April 1994, pp1-19.

[7] Hu, X., Z. Lin, A. B. Whinston., and H. Zhang, "Perceived Risk and Escrow Adoption: An Economic Analysis in Online Consumer-to-consumer Auction Markets," *The Proceedings of ICIS 2001*, 2001, pp271-274.

[8] Jarvenpaa, S. L., N. Tractinsky, and M. Vitale, "Consumer Trust in an Internet Store," *Information Technology and Management* 1(1-2), 2000, pp 45-71.

[9] Lin, Z., D. Li, B. Janamanchi, and W. Huang, "Reputation Distribution and Consumer-to-Consumer Online Auction Market Structure," *Decision Support Systems*, 2004, forthcoming.

[10] MacCrimmon, K. R. and D. A. Wehrung, *Taking Risks: The Management of Uncertainty*, Free Press, New York, 1986.

[11] March, J.G., and Z. Shapira, "Managerial Perspectives on risk and risk taking," *Management Science* 33, 1987, pp1401-1418.

[12] Quelch, J. A., and L. R. Klein, "The Internet and International Marketing," *Sloan Management Review* 37(3), 1996, pp60-75.

[13] Resnick, P., R. Zeckhauser, J. Swanson, and K. Lockwood, "The Value of Reputation on eBay: A Controlled Experiment," Working paper, 2004.

[14] Tadelis, S., "What's in a Name? Reputation as a Tradable Asset," *The American Economic Review* (89:3), June 1999, pp548-563.

APPENDIX:

A1: EBay's reputation forum before 2004

Feedback Summary

899 positives. 808 are from unique users.

3 neutrals.

1 negatives. 1 are from unique users.

[See all feedback reviews](#) for PQRtraders.

Member since: Wednesday, Jul 05, 2000 Location: United States

	Past 7 days	Past month	Past 6 mo.
Positive	7	30	88
Neutral	1	1	1
Negative	0	0	0
Total	8	31	89
Bid Retractions	0	0	0

A2: EBay's reputation forum since 2004

Member Profile: XYZ_Traders (6008) Power Seller

Feedback Score: 6008
Positive Feedback: 99.2%

Members who left a positive: 6055
Members who left a negative: 47
All positive feedback received: 7706

Recent Ratings:

	Past Month	Past 6 Months	Past 12 Months
positive	121	917	1662
neutral	1	4	10
negative	2	5	5

Bid Retractions (Past 6 months): 0

Member since: Oct-18-97
Location: United States

[ID History](#)
[Items for Sale](#)
[Learn more About Me](#)

Contact Member

Recent Trend of Ratings Business in Japan and Improving Proposal for Ratings Forecasting Model

Sakamaki, Yoshikazu

Department of Value and Decision Science Tokyo Institute of Technology Tokyo, 1528550, Japan
megacity@f3.dion.ne.jp

ABSTRACT

The change of economic circumstances in recent years has increased the gaps of credibility among firms, and the necessity to measure the credibility adequately is an important ever-increasing factor. As the result, it is essential for banks to assess the rating evaluation of borrowers appropriately in order to measure the amount of credit risk. This paper reviews preceding studies regarding model building that enables us to explain or forecast the ratings of firms issued from credit-assessment agencies. The paper then explains the features of the ordered logit model in detail that is nowadays used as a ratings evaluation model throughout the world.

Finally, some defect points of these preceding models are pointed out and the improved ordered logit model on the non-compensatory rule that enables to compensate these defect points is proposed. In addition, the appropriateness of the proposed model is verified based on the actual rating data for the companies.

1 INTRODUCTION

Banks are incorporating credit risk evaluation, as it is a very important procedure to measure the level of default probability from the point of macroscopic portfolio strategy. On the other hand, it forms the foundation of a rating system that is nowadays used for self-assessment of major banks and the new BIS Accord in financial administration.

In Japan, financial institutes that pledge to drive credit risk control forward also measure the VaR of the holding portfolio regularly and hence, the management policy is checked up based on it. However, as we verify the element of risk factors in more detail, the evaluation of individual borrowers becomes considerably important. Unless a certain level of fairness or objectivity is compensated when banking facilities evaluate the individual borrowers, it inevitably becomes difficult to decide the organic management policy based on the result of it. That means it plays a very important role in the concept of credit risk control.

When banking institutes calculate the risk-based necessary net asset in proportion to credit risk, they need to know the trend of rating change of each borrower in advance. At that time, a new BIS Accord admits two types of evaluation methods. One method is the standard method where banking institutes can use the ratings of external credit-assessment agencies, whilst the other one is an internal method in which banking institutes can use the ratings that is calculated based on its proprietary system developed by themselves.

With regards to the internal rating method, after the collapse of the 90's bubble economy, as the spread of self-assessment and discussion of a new BIS Accord, it is becoming clear that banking institutes are recognizing the internal rating system as the basic information for credit risk control.

And it becomes essential for a banking institute to assess the rating evaluation of borrowers appropriately in order to measure the amount of credit risk. Therefore, this paper introduces some preceding studies regarding model building that allows the ratings of firms issued from credit-assessment agencies to be explained or forecasted. The paper then explains the features of the ordered logit model in detail that is nowadays used as ratings evaluation model throughout the world.

Moreover, discrepancies in these models are highlighted and an improved ordered logit model is proposed that compensates for these weak points.

2 RELATED WORKS EXPLANATION OF RATINGS MODEL IN THE PRECEDING STUDY

The statistical approach for the rating of firms originated in the 1970's. In the initial study, a linear regression model was mainly used for ratings forecasting. The development of a rating model was then devised by Kaplan and Urwitz (1979). They highlighted the flawed issues contained in the preceding models based on regression analysis or discriminant analysis, and proposed an improved ratings model based on the ordered probit model for the first time. The model developed by them was very sophisticated and they succeeded to hit the rating of 44 among 64 firms without failing and moreover, even the result of ratings of the firms that could not be predicted correctly could be put at least within just one rating before/behind the correct rating.

Here, the ordered logit model is briefly explained and the ordered probit model that is nowadays used widely as ratings forecasting throughout the world is presented. At first, the ordered logit model is built by expanding the normal binominal logit model based on the explanatory variables that classifies the sample data into one of the plural statuses. It can then be used to analyze which

of the statuses the sample data can be classified only when the statuses are ordered.

Here, x_m represents the m^{th} ($m=1,2,\dots,M$) explanatory variable of firm i ($i=1,2,\dots,I$) and β_m represents the estimation parameter for x_{im} . Furthermore, we assume that financial health without an error term is given by the summation of the weighted explanatory variables. That is, we define financial health of firm i by

$$z_i = \sum_{m=1}^M b_m x_{im} \tag{1}$$

and with the latent variable z_i defined in equation (1), financial health that contains an error term is given by

$$Z_i = z_i + e_i \tag{2}$$

Moreover, we define the threshold by which the ranking of each firm is decided as

$$-\infty = t_0 > t_1 > t_2 > \dots > t_k > \dots > t_K = -\infty \tag{3}$$

Supposing the probability of firm i classified in rank s_k can be defined as p_{ik} and distribution function of the error term e_i as F , we can define p_{ik} by

$$p_{ik} = F(\tau_{k-1} - z_i) - F(\tau_k - z_i) \tag{4}$$

Here, z_i is the financial health as defined in equation (1). Moreover, paying attention to the condition of $F(\infty) = 1, F(-\infty) = 0$, we can obtain an ordered logit model if we assume logistic distribution for the distribution function F , and an ordered probit model if we assume normal distribution for F .

Based on equation (3) and (4), p_{ik} in the ordered logit model can be expressed by

$$p_{ik} = \frac{1}{1 + \exp(z_k - \tau_{i-1})} - \frac{1}{1 + \exp(z_k - \tau_i)} \tag{5}$$

On the other hand, under the condition that

$$-\infty = \tau_{-1} < (\tau_0 = 0) < \tau_1 < \tau_2 < \dots < \tau_i < \dots < \tau_{-1} = \infty$$

rating probability p_{ik} in the ordered probit model can be expressed by

$$p_{ik} = \Phi(\tau_{i-1} - \sum_{m=1}^M \beta_m x_{km}) - \Phi(\tau_{i-2} - \sum_{m=1}^M \beta_m x_{km}) \tag{6}$$

Here, F represents the cumulative distribution function of normal distribution.

Parameter estimation for these models is done as follows. That is, if we assume that the probability of firm i belonging to status s_k as p_{ik} , the likelihood function regarding parameter vector β and threshold vector t can be defined by

In this model, we assume statistical independence in financial health of each firm. Here $d_{ik} = 1$ if the financial health of i^{th} firm belongs to s_k and $d_{ik} = 0$ if the financial health of i^{th} firm does not belong to s_k . Moreover, the log-likelihood function can be defined as equation (7), and parameter vector β and t that maximizes equation (11) is estimated.

$$l(\beta, t | x) = \prod_{i=1}^I \prod_{k=1}^K \delta_{ik} \log p_{ik} \tag{7}$$

In a preceding study, Nakayama and Moridaira (1998) studied the verification of parameter estimation accuracy between the ordered logit and ordered probit models, and showed that there are no significant differences of parameter estimation accuracy between the models. On the other hand, Kobayashi (2001) proposes original statistical criteria from his own point of view and verifies that the ordered probit model can be altered as the extreme limit of multivariate probit model.

3 IMPROVEMENT PROPOSAL OF THE TRADITIONAL RATINGS MODEL

As mentioned in Chapter.2, the ordered probit/logit model is built as an expansion of the binominal logit model for classifying sample data into one of the plural statuses. These models are very useful when a certain order is given to the statuses, and it is shown in the preceding studies that these models are possible to be applied in many academic case studies. (For example, when analysts forecast the ratings of firms, credit risk and debt etc.,)

However, the ordered probit model is generally very difficult to use because of statistical reasons. Therefore, the ordered logit model is used mainly in an actual business. In this paper, we focus on the ordered logit model.

The ordered logit model is based on the scores calculated from the summation of weighted variables (linear regression analysis). That is, linearly linked variables are comprehensively evaluated and the rating in which sample data belongs to can be decided. In this sense, this type of model is built on compensatory rule. This type of model can therefore be termed as an "ordered logit model on compensatory rule" in order to distinguish it from our model.

If, in this type of model, a certain explanatory variable happens to take a huge value, the calculated probability can become far from the realistic financial health, and sample data is classified into an extremely superior/inferior rating group.

On the other hand, Katahira *et al.* (1998) have reported a very interesting study in the field of marketing research. In their study, a consumers' choice behavior model is built with the assumption that consumers make a choice behavior based on the compensatory rule. However, as we can see in such a choice behavior, no matter how cheap the merchandise is, consumers do never accept it, unless the quality satisfies a certain criteria. Consumers therefore exhibit behavior based on the non-compensatory rule throughout a variety of occasions.

Concerning this point, Dawes and Corrigan (1974), and Jhonson and Meyer (1984) have shown that under a cer-

tain condition, we can approximate a choice model on the non-compensatory rule by that on the compensatory rule. However, as pointed out in the preceding study where consumers exhibit a choice behavior based on the non-compensatory rule (Bettman and Jacoby (1976), Payne and Ragsdale (1978)), we can not approximate a choice behavior model on the non-compensatory rule by that on the compensatory rule, if a very strong negative correlation between arbitrary two variables exists (Jhonson, Meyer and Ghose (1989)).

To overcome such an issue, we will apply the concept of the choice model on the non-compensatory rule to the traditional ordered logit model. Considering the ratings forecasting model of firms, in many cases, variables used as risk factors in the model like capital adequacy ratio and interest-bearing debt have a very strong correlation with each other. Moreover, considering the relationship between the rating of firms and risk factors, there is a tendency that most of the values of risk factors go up or down according to the rating of firm changes. Taking these points into consideration, it seems to be more realistic to apply a rating forecasting model on the non-compensatory rule in which rating changes only when all or most of the risk factors considered in the model goes up/down simultaneously, rather than to apply the model based on the compensatory rule.

>From the study based on the Katahira model, Sakamaki (2003) uses the publicly disclosed financial data of firms and shows that the logit model based on the non-compensatory rule can forecast the default probability more accurately than the normal linearly linked logit model based on the compensatory rule.

When examining the preceding literature, especially in studies where a strongly negative correlation between two explanatory variables exists, it can be thought that it is very meaningful to build a statistical model on the non-compensatory rule rather than the linear model on the compensatory rule. So in this paper, taking these points into consideration, the improved ordered logit model is proposed based on the non-compensatory rule as is shown below.

Here, two types of ordered logit models are proposed based on the non-compensatory rule.

[Conjunctive Type]

In this model, it is assumed that the rating change happens only when all of the risk factors used in the model shift simultaneously to the direction of changing the ratings.

$$\begin{aligned}
 p_{ik} &= p_{ik}(\tau_{mk-1} - x_{im} > \varepsilon \geq \tau_{mk} - x_{im}, \forall m) \\
 &= \prod_{m=1}^M p_{ik}(\tau_{mk-1} - x_{im} > \varepsilon \geq \tau_{mk} - x_{im}) \\
 &= \prod_{m=1}^M \left\{ \left(\frac{1}{1 + \exp(\beta_m(x_{im} - \tau_{mk-1}))} \right) - \right. \\
 &\quad \left. \left(\frac{1}{1 + \exp(\beta_m(x_{im} - \tau_{mk}))} \right) \right\} \tag{8}
 \end{aligned}$$

[Disjunctive Type]

In this model, it is assumed that the rating change happens when some of the risk factors shift to the direction of changing the ratings.

$$\begin{aligned}
 p_{ik} &= p_{ik}(\exists m, \tau_{mk-1} - x_{im} > \varepsilon \geq \tau_{mk} - x_{im}) \\
 &= 1 - \prod_{m=1}^M p_{ik}(\tau_{mk-1} - x_{im} > \varepsilon \geq \tau_{mk} - x_{im}, \forall m) \\
 &= 1 - \prod_{m=1}^M \left(1 - \left\{ \left(\frac{1}{1 + \exp(\beta_m(x_{im} - \tau_{mk-1}))} \right) - \right. \right. \\
 &\quad \left. \left. - \left(\frac{1}{1 + \exp(\beta_m(x_{im} - \tau_{mk}))} \right) \right\} \right) \tag{9}
 \end{aligned}$$

Here, x_m represents the m^{th} ($m=1,2,\dots,M$) explanatory variable of the firm i ($i=1,2,\dots,I$). β_m represents the estimation parameter and τ_{mk} represents the threshold parameter that decides the ratings of each firm in explanatory variable x_{im} . This type of model is termed the "Ordered Logit model On Non-Compensatory Rule" in order to identify it with the normal ordered logit model. Parameter estimation for these models is done as follows. That is, if we assume that the probability of firm i belonging to status s_k as p_{ik} , the likelihood function regarding parameter vector \mathbf{B} and threshold vector \mathbf{t} can be defined by

$$L(\mathbf{Bt} | \mathbf{x}) = \prod_{i=1}^I \prod_{k=1}^K p_{ik}^{\delta_{ik}} \tag{10}$$

We assume statistical independence in financial health of each firm. Here $d_k = 1$ if the financial health of i^{th} firm belongs to s_k and $d_{ik} = 0$ if the financial health of i^{th} firm does not belong to s_k . Moreover, the log-likelihood function can be defined as equation (11), and parameter vector \mathbf{B} and \mathbf{t} that maximizes equation (11) is estimated.

$$l(\mathbf{Bt} | \mathbf{x}) = \prod_{i=1}^I \prod_{k=1}^K \delta_{ik} \log p_{ik} \tag{11}$$

Additionally, in the next chapter, applying publicly disclosed financial data to our model verifies the appropriateness of the model. Moreover, we show the proposed model is superior to the preceding ordered logit model from the view of fitness degree and forecasting power.

4 MODEL VALIDATION

In this chapter, actual data is applied to the model proposed and verifies the appropriateness of the model. The verification data used was publicly disclosed financial data issued from Rating and Investment Inc.

4-1. Outline of verification data

1) Account settlement Term

Rating data that is evaluated and published based on the account settlement data from Apr.,2000 to Mar.,2001

2) Data volume: 652 firms

(Financial data in which missing data does not contain. Moreover, in this study, the power industry is omitted because the relationship of actual management and financial health is very different with that of normal firms)

3) Financial ratio used in the model

Based on the business experience, the following six indexes that can affect the ratings of firms materially are used in the model.

1. log-transformed net asset
2. total capitalization ratio
3. ratio of fixed assets to long-term capital
4. capital adequacy ratio
5. retained earnings ratio
6. dead capacity ratio

4-2 Parameter estimation

Based on this data, parameters of the three types of models, ordered logit model based on the non-compensatory rule (conjunctive/disjunctive type), and the normal ordered logit model are estimated by maximum likelihood estimation.

Generally, the fitness of the model on the training data set is fine, but the most important question is how the model works in out-of-sample data. This is known as generalization, and there are several modeling strategies available to address it. Since the training data set is all the modeler has to work with, strategies to address generalization involve splitting the sample into sub-samples.

The simplest way to split samples is to use two sub-samples randomly selected from the full training data set. The first sub-sample is used as the actual training data set. The second sub-sample, called the *holdout* sample, is not used in the initial estimation, but later, to test the performance of the estimated model. Similar performance on both the training data set and the holdout sample is an indication that the model will perform at about the same level. If the performance of the model deteriorates

substantially on the holdout sample, the model is sensitive to specific aspects of the two sets. Such specifics could include a set of unusual observations, referred to as outliers, which are more prevalent in some data sets. Clusters of observations that show a particular set of relationships that are not always found could also cause performance differences. In estimating parameters, we have chosen 60% of the sample data at random and used for in-sample-data, and 40% for out-of-sample data.

Here, in this study, the classification of ratings is simplified based on Table 1 because there are very few number of firms classified in rating AAA, and we have to consider too many rating patterns if so-called 'notch' described by positive/negative sign is considered in the model.

Table 1. Definition of ratings used in our study

R&I Rating	Rating in this study	R&I Rating	Rating in this study
AAA AA+ AA AA-	Over AA	BB+ BB BB- B+ B B-	Under BB
A+ A A-	A	Under CCC	
BBB+ BBB BBB-	BBB		

Table 2. The number of firms contained in each rating

RATINGS	Number of firms	Ratio
AA	77	11.81%
A	201	30.83%
BBB	269	41.26%
BB	105	16.10%
Total	652	100. 00%

In this study, the parameters are estimated in each model by the best-subset selection method because it is very difficult to find the optimum combination and significant parameters effectively based on the normal variable choice method such as a stepwise algorithm. In the process of parameter estimation, twenty combinations of the initial value are generated for each parameter and the parameters are estimated by the maximum likelihood estimation method based on the initial value. In addition, the combination of parameters that minimizes the AIC and log-likelihood value are adopted as an optimum solution in each model.

The result of parameter estimation is shown in Table 3.

Table 3. Results of parameter estimates

Parameter	Normal Ordered Logit Model	t-value	Non-Compensatory Model (Conjunctive)	t-value	Non-Compensatory Model (Disjunctive)	t-value
β_a	1.782606	(16.145115)	2.304864	(10.782407)	3.102075	(7.513552)
β_b	1.703725	(1.500131)	-	-	-	-
β_c	-1.558375	(-3.405277)	-	-	-	-
β_d	-2.992869	(-2.756496)	-	-	-	-
β_e	8.349215	(8.041677)	11.123563	(10.072548)	20.005778	(8.599619)
β_f	-1.109246	(-4.757166)	-	-	-	-
T_{aa}	35.050708	(15.748833)				
T_a	31.355626	(14.957315)				
T_{bbb}	27.365669	(13.782615)				
$T_{a_{aa}}$			19.151688	(12.323935)	20.204592	(7.714131)
T_{a_a}			17.046034	(11.015636)	18.804618	(7.528707)
$T_{a_{bbb}}$			16.399593	(10.097468)	16.709588	(7.489304)
$T_{e_{aa}}$			0.650819	(7.241521)	1.064664	(0.022608)
T_{e_a}			0.478657	(9.869452)	0.187891	(7.278373)
$T_{e_{bbb}}$			0.066425	(0.851711)	0.042501	(3.267568)
Log-Likelihood	-494.490689		-467.330704		-543.979679	
AIC	1006.981379		950.661409		1103.959358	
Hitting ratio	65.48%		68.87%		59.82%	

Note: Based on the estimated parameters, the probability of a firm belonging to each rating is forecasted. If the result of forecasting and the actual rating corresponds perfectly, we regard that as a successful forecast.

The meaning of suffix x in β_x is and $T_{x,y}$ is as follows. a -> log-transformed net asset, b -> total capitalization ratio, c -> ratio of fixed assets to long-term capital, d -> capital adequacy ratio, e -> retained earnings ratio, f -> dead capacity ratio, and suffix y in $T_{x,y}$ means ratings.

5 RESULTS AND DISCUSSION

5.1 Findings

The statistical models studied in the preceding study, especially the ordered logit model has been used in the academic field of credit ratings, consumer's choice behavior etc., as a way to classify sample data into one of the ordered plural statuses. Preceding models are built based on a so-called linearly linked ordered logit model based on the compensatory rule, in which weighted explanatory variables used in the model are linearly linked. As mentioned in the preceding chapter, in the linear model based on the compensatory rule, if there exists a very serious relationship between two particular variables such as capital adequacy ratio and interest-bearing debt, the forecasting power decreases. Moreover, it is likely to happen that if one of the risk factors used in the model has extremely large/small values, a wrong rating evaluation inconsistent with the actual management status will be given to the firms.

Paying attention to each risk factor in the preceding model, it was difficult to know the threshold value among the ratings by each factor. Considering these points, an ordered logit model is proposed based on the

non-compensatory rule in order to overcome these issues in this paper, and the appropriateness of the model is verified by applying actual data to it.

It can be shown that new model proposed in this study can improve all of the log-likelihood, AIC criteria, and the hitting ratio of rating forecasting compared with the ordered logit model. Moreover, regarding the risk factor used in the model, we used all of the risk factors available in the model to optimize the fitness in the preceding model. Whilst on the other hand, it could be shown that we could get a higher fitting degree by using only two kinds of risk factors in the proposed model. The reason for this can be thought to be that in the preceding model, many kinds of risk factors that have a strong correlation to each other are contained in the model. This brings an adverse affect to the model, which as a result, leads to a declined forecasting power.

Furthermore, the result of parameter estimation is considered. Comparing the result of parameters estimated in each model, whilst regarding the identical parameter, the size of each parameter seems to vary among the models. However, the positive/negative sign of the identical parameter matches perfectly. It can therefore be assumed that the parameter estimation was performed accurately.

5-2 Future Investigations

Finally, there are issues of this study that require improvement for future work. Generally, when we estimate the parameters in the normal logit model, we can estimate the parameters effectively by using a parameter estimation algorithm such as a stepwise method. However, in this proposed model, the following point still remains where we have to assume two kinds of parameters for one explanatory variable, so that it is impossible to apply the normal stepwise algorithm to the non-compensatory type of model.

This study has adopted the best-subset selection method as the parameter estimation algorithm. However, this method does not always choose the optimal combination patterns of parameters of the model. Moreover, the number of combination patterns of variables increases rapidly and the length of time needed to find the optimum value of the parameters increases as the number of variables used in the model increases.

Regarding our model on the non-compensatory rule, applicability of the model is reported to be very various in the preceding study. For example, it can be used not only for ratings model but also for consumer's choice behavior model in marketing research and traveler's decision making for transformation method in social engineering science, and so on. Academic fields to which statistical model on non-compensatory rule can be applied, should spread rapidly in the near future. Variable choice algorithms such as stepwise, backward, and forward methods are proposed in the study of a linear model based on the compensatory rule. We expect the same type of variable choice algorithm that enables the analysts to use the model easily should be developed in the future.

REFERENCES

- [1] Bettman, James R. and Jacob Jacoby (1976), "Patterns of Processing in Consumer Information Acquisition", *Advances in Consumer Research*, 3, 315-320
- [2] Dawes, Robin M and Bernard Corrigan (1974), "Linear Models in Decision Making", *Psychological Bulletin*, 81 (March), 95-106
- [3] Johnson, Eric.J. and Robert J. Meyer (1984), "Compensatory Choice Models of Noncompensatory Processes: The Effect of Varying Context", *Journal of Consumer Research*, Vol. 11, June, pp.528-541

- [4] Johnson, Eric.J., Robert Meyer and Sanjoy Ghose (1989), "When Choice Models Fail: Compensatory Models in Negatively Correlated Environments", *Journal of Marketing Research*, 26(August), 255-270
- [5] Kaplan, R. S. and G. Urwitz, (1979), "Statistical Models of Bond Ratings: A Methodological Inquiry," *Journal of Business*, 52, 231-261, 1979.
- [6] Payne, John W. and E.K. Easton Ragsdale (1978), "Verbal Protocols and Direct Observation of Supermarket Shopping Behavior: Some Findings and a Discussion of Methods", in *Advances in Consumer Research*, 5, 571-577
- [7] Kobayashi, Masato (2001), "Junjo Probit Model No Test To Syasai Kakuduke Data He No Ouyou [Test of Ordered Probit Model and Application for Bond Ratings Data]" *Monetary and Economic Studies*, April, 2001, Institute for Monetary and Economic Studies, Bank of Japan
- [8] Katahira, Hideki (1998), "Logit Bunseki Wo Mochiita Manzoku-Ka Model [Satisfaction Model with use of Logit Analysis]", *Syohi-Sya Koudou-No New Directions [New Directions of Consumer's Choice Behavior]*, Kansai-Gakuin-Daigaku [Kansai-Gakuin University]
- [9] Sakamaki, Yoshikazu (2003), "Hi-Hosyogata Model-Wo Mochiita Default Kakuritsu Yosoku Model-No Kaizen Teian [Improving Proposal for default probability forecasting model with use of non-compensatory rule]", *Collected Resume for Convention, 2003, Summer, The Japanese Association of Financial Econometrics & Engineering*, 154-170
- [10] Nakayama, Megumi and Moridaira, Souichiro (1998), "Kakuduke Sentaku -Kakuritsu No Suitei To Sinyou Risk-Ryo [Forecasting of Rating Choice Probability and Credit Risk]", *Collected Resume for Convention, 1998, Summer, The Japanese Association of Financial Econometrics & Engineering*, 210-225

AUTHORS BIOGRAPHIES

Yoshikazu, Sakamaki is a student of doctor course in Tokyo Institute of Technology, Department of Value and Decision Science, Japan. He received a B.Eng in the Department of Control and System Engineering from Tokyo Institute of Technology and his M.Eng in the Department of Industrial Engineering and Management from Tokyo Institute of Technology. His recent work has involved financial engineering, marketing research with use of statistical approach. His email address is megacity@f3.dion.ne.jp.

Security Alert Management in E-Business Networks

Allan Lam, Pradeep K. Ray

School of Information Systems, Technology and Management,
University of New South Wales, Sydney, Australia
p.ray@unsw.edu.au

ABSTRACT

Security management has become a major concern in today's e-business systems due to ever-increasing attacks on enterprise servers. This has led to the increasing sophistication of network security tools and systems in e-business networks that involve a number of organizational entities cooperating over computer communication networks. Many large organizations are outsourcing the management of e-business networks. This paper examines the problem of security management in the context of an Management Service Provider (an organization that provides remote management of e-business networks). Existing security tools (e.g., Intrusion Detection Systems (IDS)) assist us in detecting attempts by unauthorized users to get access to networked information resources. However, the management of IDSs offers some interesting challenges (e.g., false alerts). This paper presents a policy-based management framework to solve this problem.

Keywords: Outsourced IT management, IDS Alerts, Management Service Provider

1. INTRODUCTION

E-business systems allow organizations to focus on their core strengths, while outsourcing many aspects of the business to business partners in a value chain. For example, many e-businesses now outsource the management of their IT networks to specialist organizations [7]. The increasing ferocity of cyber-attacks is making organizations think seriously of more and more sophisticated security management solutions. Organizations now have firewalls, anti-virus, authentication, and Intrusion Detection Systems (IDS). However, many of these facilities are either unused, or under-utilised due to problems, such as false alarms in IDS [1]. This paper presents a security management strategy for large e-business networks from the perspective of outsourced management.

The paper is organized as follows. The Section 2 starts with a background of current intrusion detection systems and the need for frequent management interventions. This is followed in Section 3 by our proposed architecture for Security Alert Management (based on management policies) that helps reduce frequent false alarms. Section 4 presents a case study of the implementation of this architecture in a large organization providing management services to e-business networks in the Asia-pacific region. Finally, Section 5 concludes the paper.

2. BACKGROUND

Many network security technologies require regular human intervention in order to receive the best protection. Consider firewalls and intrusion-detection systems (IDSs). Typically, these devices are setup so that they log information to a database. It is then up to the security engineer to analyse these logs and to identify network attacks or suspicious behaviour if, or

while, they are occurring. Unfortunately, this can be very time consuming.

Logs are generated and typically contain information for the following three events:

A packet that is allowed to pass through a firewall.

A packet that is blocked (denied access) at a firewall.

A packet that satisfies one of the signatures in the IDS database.

Thresholding rules are then applied to the above events to generate security alerts. Unfortunately, because they work on data from firewalls and IDSs, they also inherit the major problem with them, e.g., the generation of numerous false positives [2].

This research reported in this paper aims to develop a framework whereby security alerts, in particular false or routine alerts, can be analysed and responded to automatically, saving precious employee time and the employer's money. Routine alerts are those that we wish to appear but which we do not want to analyse. False and routine alerts are composed of many events. Our proposed framework is called Security Alert Management System (SAMS). This is a Policy-based management framework as discussed in [4], [5] and [6].

3. SECURITY ALERT MANAGEMENT SYSTEM (SAMS)

SAMS uses the term policy to refer to a particular rule that defines a false or routine alert. Every policy is instantiated from a policy type. A policy type represents a set of policies that possess a similar underlying idea/pattern. For example, consider the following policy types:

- If pattern A appears at least Z% of the time in the alert, then acknowledge this false/routine alert.
- If pattern A appears at least Z% of the time in the

alert, AND if a pattern B appears at least Y% of the time in the alert, then acknowledge this false/routine alert.

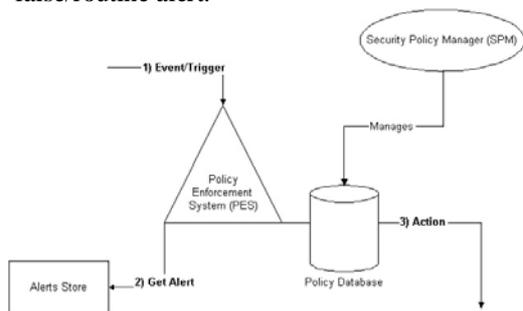


Figure 1: SAMS framework

The above policy types can be used to instantiate the following policies:

- If the source IP address 127.0.0.1 appears 90% of the time in the alert, then acknowledge this false/routine alert.
- If the source IP address 127.0.0.1 appears 40% of the time in the alert, AND the destination IP address 234.22.12.120, then acknowledge this false/routine alert.

The policy types then really have no constraints, other than their realisation into code, which will be discussed later.

3.1 SAMS Framework

The framework presented in Figure 3 provides a design for the development of a system that analyses security alerts. It is called the Security Alert Management System (SAMS) framework. At a high level of detail, SAMS will wait for an event or trigger to start the alert analysis process. Once this event has been received, the alerts are retrieved from an alerts store. Each alert is then analysed with its corresponding policy, which is determined by the domain that the alert belongs to. Each policy will determine whether or not the alert is a false alert or a routine alert. The policy will also determine what action to carry out based on that analysis.

The simple framework of SAMS reflects the nature of the problem, which is to analyse and respond to security alerts. However, its simplicity hides the powerful inner workings of the policies. The more verbose or complex your false or routine alerts, the more complicated your policies become. As shown in Figure 1, the framework is made up of three main components, the Policy Enforcement System (PES), the Policy Database and the Security Policy Manager (SPM). These will be described in the following three sections.

3.2 Policy Enforcement System (PES)

The PES does three things, as shown in Figure 2 with a

triangle:

A) Intercept an event/trigger. This event/trigger can be a mouse click or a timed event and invokes the PES process.

B) When the PES is invoked, alerts are read from whatever format that they are stored as.

C) Each alert, based on its domain is then compared to its corresponding policies found in the policy database. If a policy is found that matches the alert then the action for that alert is automatically carried out.

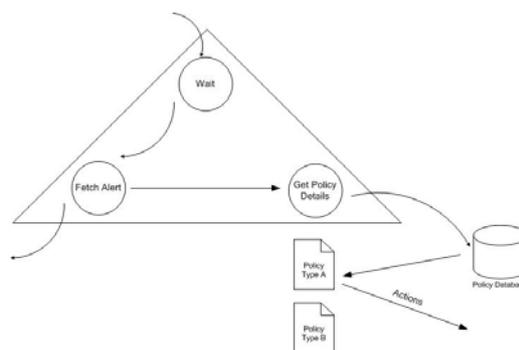


Figure 2: PES process

The PES ‘waits’ for the event/trigger to start the process. Upon receiving this, it will ‘fetch alerts’ from the alerts store. Finally, it will ‘get the policies’ for this alert from the policy database. The policies that it retrieves are based on the alert domain. The policies are compared to the alerts and if one is found to “match”, then the appropriate ‘actions’ are carried out. The last step is actually a little more involved, which will now be discussed. Recall from the beginning of section 3 that we wish to define different policy types. Each policy type will define a false or routine alert. The PES will actually store the policy type definition, represented as ‘Policy Type A’ and ‘Policy Type B’ in the above diagram. The policy database will then store the particulars for the policy type definitions. For example, the PES may have the following as ‘Policy Type A’:

If pattern A appears at least Z% of the time in the alert, then acknowledge this false/routine alert.

The policy database may then have a table such as the following:

alert_domain	pattern_A	percentage_Z
1	Shellcode.*	90

So if an alert that belongs to the alert_domain 1 is encountered, then the corresponding pattern (Shellcode.*) and percentage (90) will be replaced in the policy type. Encountering such an alert will apply this policy to it:

If pattern Shellcode.* appears at least 90% of the time in the alert, then acknowledge this false/routine alert.

If the alert satisfies this policy, then it will be

acknowledged. Note that the pattern supports regular expressions. By separating the policy type and its specific details, we allow a policy to be applied to various domains and alerts, without producing repetitive policies. As the policy types need to be coded, this also reduces code redundancy.

In summary, the PES has been designed so that it stores the programming logic to analyse alerts. It in fact contains the skeleton of the logic. The remainder of the programming logic is obtained from the policy database. The reason for this separation is that many alerts have similar analysis. For example, many alerts are analysed based on what event occurs the most often. So, instead of replicating the code for this analysis across all domains, we can save disk space and redundancy by storing only the skeleton of the code and storing the finer details in the database. Creating new policy types therefore involve coding them into the PES.

PES is the only component that integrates with a current thresholding system.

3.3 Policy Database

This represents the database of policies. As described in the previous section, it will store the details of the policies in a policy table. Each domain is also represented in the Policy Database. This only makes sense since it is natural and easy to link domains with their particular policies via table joins. A schema such as the following can be employed for security alert management systems developed from SAMS (Figure 3).

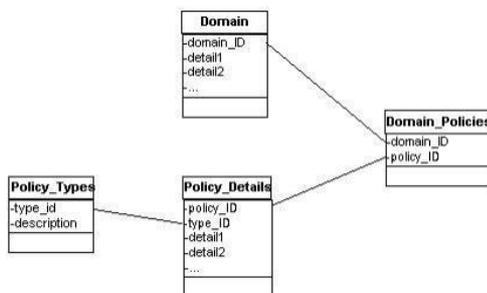


Figure 3: Policy Database schema

The table Domain will contain information about that particular domain. Policy_Details will contain information about all policies that have been created for the system and Domain_Policies is a lookup table matching the policies to the particular domains. So to retrieve the details for a particular policy the following query might be performed:

```
select p.detail1, p.detail2 from domain_policies d,
policy_details p where d.domain_ID=1 and
d.policy_ID=p.policy_ID.
```

The attributes returned from the above query can then be substituted into the policy type definition defined in

the PES. The Policy_Types table helps with this task by associating a policy instance in the table Policy_Details to the policy type definition in the PES.

3.4 Security Policy Manager

The SPM is the GUI management tool that will allow you to manage the Policy Database. It will allow you to enable, disable or edit policies for your defined domains. The SPM will present the information from the database in an easy to read format allowing rapid creation of the policies.

As discussed in the beginning of section , policies are instantiated from policy types. Hence, when we enable a policy, we are actually creating an instance of a policy type. Every instance is created by inserting the details for the policy into the policy database, discussed in the previous section. The policy created is then applied to a selected domain.

When we disable a policy, we are actually removing an instance of the policy from the policy database. This involves removing various rows from the policy database to remove both the policy instance and its association with a particular domain. Note that when we disable a policy, we do not actually remove the policy type definition from the PES. We are only removing an instance of it, which is defined in the policy database.

Editing a policy allows us to edit the details of a policy instance. For example, consider the following policy: If pattern Shellcode.* appears at least 90% of the time in the alert, then acknowledge this false/routine alert. The user could modify this policy by changing the pattern or changing the percentage value. Note also that editing a policy does not edit the policy type definition. Editing a policy only edits a policy instance. Editing a policy will perform a database 'update' query on the policy_details table.

Note that whenever a new policy type is created in the PES, the SPM must also be updated so that it recognises it. The SPM must be updated so that one can enable, disable and edit policies based on the new policy type.

4. CASE STUDY

This section will use the SAMS framework to develop a working system. The system will be used on a thresholding system already in place at a company X. The company has asked its name to be withheld. The company is a leading IP networking and communications solutions provider in the Asia-Pacific region.

Of interest to this paper is the organisation's security division. The security engineers, among other tasks,

monitor the alerts from various firewalls and IDSs for various customers. The security division has implemented a thresholding system, which analyses events from firewalls and IDSs and based on predefined thresholding levels will generate alerts. These alerts are displayed to security engineers via an internal intranet web page.

The SAMS framework helps design a system that automatically acknowledges false and routine alerts produced from the thresholding system. According to the engineer, on average it takes one minute to analyse one security alert. On an average day for the organisation's security division, they receive 300 alerts. This means that every day, 300 minutes is used to analyse false alerts. If there are 230 working days per year, that equates to 1150 hours or 143 working days that are spent on analysing alerts.

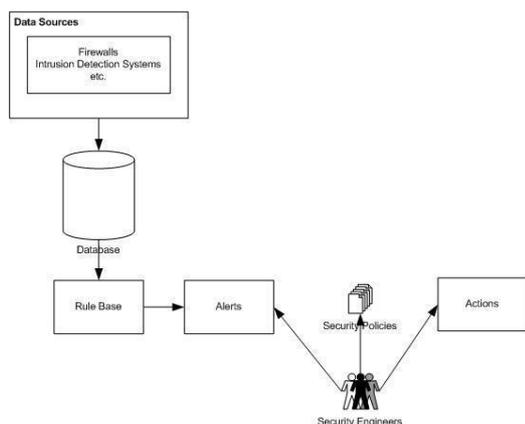


Figure 4: Current security architecture of company X

4.1 Current Security Architecture

Before we design a SAMS system, we must first analyse the current architecture to see where it fits in. The following diagram shows the current data flow of the organisation's security division:

The data sources log their events to a PostgreSQL database. The thresholding system rule base then generates the alerts. Security engineers will analyse these alerts and perform the required actions. As indicated Figure 4, the security engineers are actually using policies when they are analysing the alerts. That is, the security engineers are applying certain rules for the analysis that they carry out.

4.2 Design

This Section presents the design of the proposed system based on SAMS (Figure 5). However, we first need to identify how our alerts are being stored.

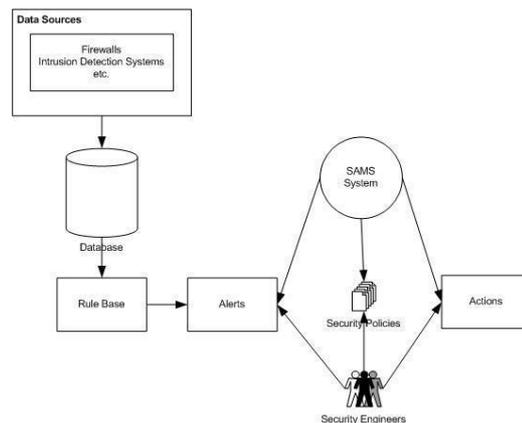


Figure 5: Proposed SAMS Architecture

The PES will therefore perform a similar task. It will read in the alerts from a HTML file and In the case of the organization's thresholding system, the alerts are stored as HTML files. Security engineers analyze alerts by entering an intranet web site. This web site provides the alerts and the events that make up in the alerts in a tabular format. It is then up to the security engineer to analyze these and acknowledge that they have been analysed. PES will only acknowledge the alerts if the policies in its database match the alerts.

The next step is to identify the domains. In the case of thresholding system, these will be the customers. Since the organisation manages customer networks, it is only logical that we distribute policies according to this domain. The existing database already has this information in a table called 'customers'. The schema for this table is as follows:

```

customers(id, name, fullname, heat_id, gnc_name, speech_name)
    
```

Hence for our SAMS system we will need to create these tables:

```

policy_details(policy_id, description, ...)
cust_policies(cust_id, policy_id)
policy_types(type_id, description)
    
```

The full schema of policy_details will be worked out later.

Our final step is to develop the SPM. Because the existing thresholding system is web-based, the SPM designed here will also be web-based. Similarly, because the existing thresholding system is developed in Perl, our SAMS system will also be developed in Perl.

4.3 Implementation

After discussion with the organisation's security staff, it was determined that the main policy type that should be created first, was one that specified a pattern that was most prevalent. That is, in the analysis of every alert, false and routine alerts tended to have a pattern whereby certain elements would be repeated. As an example of such an alert, refer to the following that has been used throughout this report:

- If pattern Shellcode.* appears at least 90% of the

time in the alert, then acknowledge this false/routine alert.

This is the first policy type that was coded into PES and which the SPM will recognise. Hence, after this decision was made, the schema for the table policy_details is as follows:

policy_details(policy_id, devicetype_id, type_id, details)

The important attributes are described here:

- type_id: this is a foreign key to the policy_types table and associates a policy with its policy type.
- details: this will be a string that contains the details for a policy. Since we are implementing the above policy type, this string will contain the pattern to look out for and the percentage that this pattern appears in the alert. More precisely, it will contain what specific details of an alert contain the pattern, what the pattern is and the percentage value.

When the user is creating a policy, they must first select which components of an alert make up the “key”. The key is a combination of the attributes in the above table that the policy type definition will concentrate on when it is analysing it.

The SPM was then designed so that it could properly manipulate the policy database. As mentioned, all development work for this system was done in Perl. The existing thresholding system was developed in Perl, so for compatibility issues, it was logical to select this as the development technology.

4.4 Results

A successful SAMS system was developed on the development network. The following table shows how many alerts could have been automatically acknowledged if the SAMS system was put into production.

Time (for 26 November 2003)	Number of alerts that could have been automatically acknowledged
00:00 to 01:00	6
01:00 to 02:00	3
02:00 to 03:00	3
03:00 to 04:00	3

This is but a four-hour sample of the number of alerts that could have been automatically acknowledged. It would equate to 15 minutes of saved alert analysis time. For another random sample:

Time (for 27 November 2003)	Number of alerts that could have been automatically acknowledged
13:00 to 14:00	7
14:00 to 15:00	9
15:00 to 16:00	5
16:00 to 17:00	6

The above would have saved 27 minutes of alert time. These alerts are all based on the same policy type specified in the previous section. Due to the nature of networks it is hard to predict how many alerts on average could be automatically acknowledged with this policy type. But it is clear from the above samples that significant time can in fact be saved.

Future work will involve the creation of policy chains, introduction of policy priorities and the development of a policy type creator, since with the current framework, it is quite cumbersome to add other policy types. Also it is necessary to integrate the SAMS framework with the network and systems management framework [3].

5. CONCLUSION

In this paper, we have proposed a new framework, called SAMS, for the analysis of security alerts in e-business networks. This framework incorporates foundational concepts in policy-based management to solve the problem of false alerts that threatens to render many security tools (e.g., intrusion detection systems) useless.

The framework that is proposed in this paper is free from many constraints on the policy definitions, unlike in the case of existing management policy frameworks. In fact, policies defined in this framework are constrained by the author’s ability to code them. This is hopefully a first step towards a framework that may become universally used in any system that generates alerts.

The time and money wasted in analysing false and routine alerts is great as proven by our case study figures. A system developed from SAMS can also be viewed as an “engineer-in-a-box”. Because the security engineer is telling the SAMS system what constitutes false/routine alerts, the system is really acting like a security engineer. As a consequence, if the security engineer decides to leave the organization for which he/she works, then the knowledge base of the engineer will be retained.

Such policy-based management frameworks are likely to be used more increasingly for security management in future.

ACKNOWLEDGEMENT

This work was partially supported by the Australian

Research Council (ARC) SPIRT Grant #C00107103.

REFERENCES

1. Axelsson, S., "Intrusion Detection Systems: A Survey and Taxonomy", Technical Report of the Department of Computer Engineering, Chalmers University of Technology, Sweden, March 2000
2. Blackman, D., "Intrusion Detection is failing: Enter Intrusion Management", July 2002, <http://www.itsecurity.com/papers/pentasafe1.htm>
3. Distributed Management Task Force, Inc. (DMTF), "Common Information Model (CIM) Specification", version 2.2, June 14, 1999, <http://www.dmtf.org/spec/cims.html>
4. N. Dulay et al. A Policy Deployment Model for the Ponder Language (2001)
5. B. Moore, J. Strassner, E. Elleson, "Policy Core Information Model -- Version 1 Specification", Feb 2001, <ftp://ftp.rfc-editor.org/in-notes/rfc3060.txt>
6. Proctor, P., "The Practical Intrusion Detection Handbook", Prentice Hall, 2001
7. Ray, P., "Integrated Management from e-Business Perspective: Concepts, Architectures and Methodologies", Kluwer Academic/ Plenum Publishers, January 2003
8. P. Yarang, P. Ray and D. Maher, "Profiling Cyber Attacks using Alert Regression Profiles", IEEE Globecom2003 Symposium on Network Security, San Francisco, Nov 2003

The Impact of a 'Theological' Web Interface on Trust in the Context of Electronic Commerce

Haytham Siala¹, Feroz Siddiqui²

¹KIT eLearning, an Affiliate of University of Liverpool (UK),
Arena Boulevard 61 –75, 1101 DL Amsterdam Z.O. The Netherlands
Haytham.Siala@embanet.com

²Department of Information Systems and Computing, Brunel University,
Uxbridge, Middlesex, UB8 3PH
Feroz.Siddiqui@brunel.ac.uk

ABSTRACT

Researchers and academics from diverse disciplines have highlighted the role of 'trust' for establishing and strengthening existing relationships between individuals and organisations in the commercial and the social context. Various models have been proposed which aim to identify both the antecedents and outcomes of trust displayed towards e-commerce Web sites. Increased trust is generally shown to increase positive user attitude, which in turn is believed to increase the intention to buy. Studies have shown the antecedents of trust include variables such as the perceived reputation and size of the vendor organisation. The current paper investigates the role of trust on e-consumers from a cultural perspective. Religious sub-cultures have been targeted as the main study area for this paper's research investigation. Participants recruited from Christian, Muslim and other faiths were asked to interact with online bookstores that identified with a Christian, Muslim or Neutral orientation. Neutral Web sites were those sites classified as devoid of any religious marketing, branding and logos. Trust and attitudes towards the Web sites were measured and this data was used to test the hypothesis that same-religion sites would be found more trustworthy and appealing than other religion or neutral sites. This hypothesis was partially supported, but only for the Muslim participants. It was found that the Muslim group expressed significantly more trust in the Muslim site compared to the Christian site. They also expressed significantly more positive attitudes towards the Muslim online bookstore than the other two sites. The implications of these results for theories of web based trust and attitude are discussed along with the practical implications of the findings.

Keywords: Web-based Trust, Web-based Attitude, Culture, Religion

1. INTRODUCTION

The continuous growth in internet shopping has incited researchers to investigate the various factors that influence consumer purchase decisions in the web context. Trust is one of the variables which is receiving considerable attention. The lack of trust was often cited as the significant barrier to engage in e-commerce transactions [6,8]. Studies such as Jarvenpaa et al (2000) substantiate the supposition that there is a relationship between trust in an online store and willingness to buy. Other studies have also suggested a number of HCI and organizational factors that could contribute to increased trust in a Web site [6,16,18]. This paper considers the nature of the relationship between culture and trust in the e-commerce context. It asks whether a consumer's cultural affiliation can itself act as an antecedent to trust. The main theme of this study tests the hypothesis that participants will trust and like an e-commerce site affiliated to their own religion more than one affiliated to a different religion or to no religion.

2. TRUST IN THE E-COMMERCE CONTEXT

Definitions of trust vary from one context to another. Husted (1989) argues that trust is complicated due to the wide spectrum of approaches that define and describe it.

In general the concept of trust involves a relationship between the trustor and the trustee [12]. A precondition for trust to manifest and become a relevant concept is the presence of risk or uncertainties in a potential decision [10]. Internet purchasing certainly meets this criterion since all transactions in e-Commerce are dealt with remotely. Our research involves studying trust from a one-to-many relationships perspective (for example between an individual and an organisation) since it is this type of trust which an individual consumer holds in a Web-based retail organisation. In recent years a number of researchers have begun to investigate the role of consumer trust in e-commerce purchase decision making and several models have been developed [6,16,19,31]. In such models trust is generally seen as a factor which can affect a consumer's willingness to buy [31]. Jarvenpaa et al (2000) also proposes that this relationship is mediated by attitude (a positive relationship) and risk perception (a negative relationship). An empirical study carried out in a number of different countries corroborates this model [15]. Some researchers focused on identifying the antecedents of trust in e-commerce stores. Kim and Moon (1998) provide empirical results which suggest that the visual components of a Web site (such as color and graphics) can be manipulated to induce trust in Web visitors [18]. Egger's (2002) Model of Trust for

Electronic Commerce (MoTEC) suggests that pre-purchase knowledge (including store reputation), interface properties and informational content (for instance information provided by the store about their policies/guarantees) are all trust-instigating factors [6].

3. TRUST AND CULTURE

There are several reasons to assume that culture may be an important factor in on-line trust. Some studies found the ingrained cultural values that form an essential part of the cultural heritage of a person to be influential in evoking an individual's trust [9, 29]. In general, trust in its cultural context is credited with being the social 'glue' for sustaining bonds between members of cultural groups [32,33]. Individuals are more likely to perceive members outside their social circles as more untrustworthy than their in-group members [2].

In a collectivist culture, an individual is regarded as a part of the group that s/he affiliates [9]. Collectivists' behaviour is usually regulated by group norms. Credit or blame is a 'group' responsibility and strong distinctions are made between in-group and out-group members. This is in contrast to individualist cultures where the distinction between in-group and out-group members often overlaps or is altogether non-existent [36]. Further, collectivists value homogeneity within in-groups [32]. Thus, they are more willing to trust individuals and sub-groups within the same culture (in-group trust). In this research, we have chosen to explore the role of membership in a collectivist culture in the formation of trust in online stores. As has been mentioned earlier, members of a collectivist culture assign a higher level of trust to others from the same in-group than to those from outside. The question we address is whether this effect also applies to interactions with an interface, particularly a Web-based electronic commerce interface. In order to explore the proposed relationship between trust and culture, we needed to identify participants from collectivist cultures and have chosen religious identification to classify our cultural in-groups (religious sub-cultures). According to Iannaccone (1995), Islam, Judaism and Christianity share the traits of a collectivist culture. Devout followers of such religions are expected to abide by the rules set by their religious doctrines, for example, being strictly committed to the religious practices and membership of the group [13,27,28]. Membership of a collectivist religion thus provides a good example of where in-group trust should be present. We have selected the collectivist religions of Christianity and Islam in our research. Using two collectivist religions enables us to test the generalizability of any effects found.

3.1 Research hypotheses

On the basis of the discussion above we propose the following research hypotheses:

H1: Members of a collectivist religion will trust a same-religion web-site more than a neutral web-site or other religion web-site.

This hypothesis can be rephrased more specifically according to the religious groups used in this study, thus:

H1a: Members of the Muslim religion will trust a Muslim web-site more than a neutral web-site or other religion web-site.

H1b: Members of the Christian religion will trust a Christian web-site more than a neutral web-site or other religion web-site.

H2: Members of a collectivist religion will express more positive attitudes towards a same-religion web-site than a neutral web-site or other religion web-site.

Again this hypothesis can be rephrased more specifically as:

H2a: Members of the Muslim religion will express more positive attitudes towards a Muslim web-site than a neutral web-site or other religion web-site.

H2b: Members of the Christian religion will express more positive attitudes towards a Christian web-site than a neutral web-site or other religion web-site.

Note that H2 follows from H1, since theories of trust suggest that trust is related to attitude. The following sections describe the research which was followed in order to investigate these hypotheses. Note that the scope of this research is confined to initial trust attitude formation.

4. EMPIRICAL WORK

An empirical study with 91 participants was conducted to explore the relationship between religious identification and trust in on-line stores. The experimental task involved participants interacting with religious or neutral e-commerce book stores. Questionnaire responses were used to measure the main dependent variables of web-based trust and web-based attitude. A separate questionnaire was used to measure demographic characteristics of the sample together with the extent of their religious affiliation and web experience.

4.1 Construct measurement

A number of measures were developed and used in this research. Where measures were developed from scratch, we followed the procedure recommended by DeVellis (1991) and Bearden and Netemeyer (1999) for the development of psychological and latent construct measures which involved piloting an initial 'pool' of items drawn from the literature that are meant to represent the constructs. The Likert and categorical scaling procedures were used for developing the constituent items of the constructs following the guidelines provided by Zikmund (1977). To counter for

the response set of acquiescence, we reversed the Likert scale rating and the direction of the wording of some items [5]. The resulting items were tested in a pilot trial (involving 50 participants) and some items were reworded to remove ambiguities identified during this trial. It had been argued that the convergent loadings of items on a separate factor is a trait of construct validity [1,5] and that in general, factors extracted from a factor

analysis are assumed to represent the validity of the underlying latent constructs being investigated in a research study [5]. We have therefore performed a factor analysis on the constituent items representing the different constructs to see if they are valid. Cronbach's Alpha is used as a measure of the internal consistency of each scale used. A Cronbach's Alpha values of 0.80 or more are generally required for widely used scales [14]

TABLE 1. Characteristics of the sample cross-tabulated against religion (N = 91).

	Christian (%)	Muslim (%)	Others (%)	Total within Sample (%)
Gender				
Male	62	79	50	65.9
Female	38	21	50	34
Ethnicity				
Arab	0	29	0	12.1
Asian (Other)	3	8	17	8.8
Bangladeshi	0	3	0	1.1
Black African	3	3	0	2.2
Caucasian	3	0	0	1.1
Chinese	0	0	25	6.6
European/White	83	8	21	35.2
Hispanic	3	0	0	1.1
Indian	0	3	34	9.9
Kurdish	0	3	0	1.1
Mauritian	3	0	0	1.1
Pakistani	0	39	0	16.5
Persian	0	5	0	2.2
Srilankan	0	0	4	1.1
Usage of Internet				
< 6 months	0	0	0	0
1 year	10	3	4	5.5
2-3 years	28	39	30	33
> 3 years	62	58	68	61.6

4.1.1 Dependent variables

Measures of web-based trust and web-based attitude were developed following the procedure described above. Several literature sources from different disciplines were consulted in the development of the constituent items of web-based trust [3,4,7,17,21,23,25,34] and Web-based attitude [11,22,24]. Following the pilot study a final version of the questionnaire was produced with items representing 'web-based trust', 'web-based attitude' and the 'intention to buy' from an online store.

The web-based trust (Cronbach's Alpha = 0.88) and web-based attitude (Cronbach's Alpha = 0.93) constructs were both reliable. The Intention-to-buy

loaded on the web-based attitude factor and was therefore included within this general attitude measure. There was a significant correlation ($p < 0.05$) between the scores on the measures of web-based trust and web-based attitude which lends ground to models that posit on-line trust to be an antecedent to attitude and willingness to buy [16].

4.1.2 Demographic variables

Several categorical demographic measures were used including one for the independent variable of religious identification (Christian, Muslim, none-or-other), as well as gender and ethnicity. In addition a number of demographic scales were used or developed to measure variables that might be expected to be important in the interpretation of the results (e.g. experience in using computers and the internet). We also wanted to establish the extent to which culture affects their general purchasing habits specifically the extent to which participants were inclined to buy from vendors within their own religious group. To measure this construct we adapted the pre-existing consumer ethnocentrism scale or 'CETSCALE' [26] so that it referred to religion rather than ethnicity. We refer to our new measure as the consumer religio-centrism scale (RELSCALE). The validity of the scales was checked through factor analysis and after some modifications the religious commitment scale (Cronbach's Alpha = 0.92) and the RELSCALE (Cronbach's alpha = 0.86) were both found to be reliable.

TABLE 2: Mean scores on the religious commitment scale and consumer religio-centrism scale.

Religious identification	Religious Commitment Mean (standard deviation)	Consumer Religio-Centrism Mean (standard deviation)
Christian	2.72 (0.88)	1.39 (0.52)
Muslim	4.11 (0.58)	2.83 (0.77)
None-or-Other Religion	2.15 (0.80)	1.40 (0.42)

4.2 Method

4.2.1 Experimental design

A pseudo-experimental approach was used in the empirical study. A 3 x 3 mixed-factorial design was used. The naturally occurring variable of religious identification of the participants was used as the between subjects independent variable with three levels: Christian, Muslim and None-or-Other. The religious affiliation displayed by an e-commerce web site was used as the within-subjects independent variable, again with above three levels. The dependent variables were web-based trust and web-based attitude.

We decided to use online bookstores as our 'testbed' sites since this category of purchase has been the focus of a number of other studies [15]. We were also able to find web sites within this category which were clearly affiliated to particular religions. The use of real web sites increases the ecological validity of the results.

Our hypotheses were directly tested through the inclusion of the Christian and Muslim groups of participants. However, since we were not able to control for all of the features of the web sites used, we also included a control group of participants, not affiliated to either of the target religions. This allowed us a means to check for the role of extraneous (non-cultural) variables. If the pattern of responses for all three groups across the three web sites were similar it would suggest that the results could be explained by non-cultural factors alone. If they vary by religious identification for the religious groups, and this pattern differs from that observed for the non-affiliated participants, we have stronger grounds for attributing the findings to the effects of culture.

4.2.2 Participants

Participants in this study were recruited from within the student population at Brunel University, West London, UK. This population was chosen since it is culturally diverse and generally experienced in using the web.

Ninety one participants took part in the study. Among them, 29 were categorised as Christian, 38 as Muslim and 24 as None-or-Other religion. Their demographic characteristics and web experience are summarised in Table 1 (see previous page).

The data indicate that the majority of participants within all religious categories are experienced web users.

However, there are large gender differences between the religious groups; nearly 80% of the Muslim sample are male compared to 62% in the Christian sample and 50% in the no-or-other religion group.

Mean scores on the religious commitment scale and consumer religio-centrism scale are shown in Table 2. The Muslim participants in our sample appear to be more strongly committed to their religion than the participants in the other groups, they also appear to be more consumer religio-centric than participants in the other groups. These differences were tested using the Krustal-Wallis test (this non-parametric alternative to one-way ANOVA was used since the assumption of homogeneity of variance was not met for these data sets). It was found that religious commitment varied with religious identification in our sample (Chi-Square(2) = 52.2, $p < 0.001$). Post hoc Mann Whitney tests confirmed that there are significant differences between the Muslim participants and Christian participants ($U=111$, $p < 0.001$), and between the Muslim and the None-or-Other religion participants ($U=24.5$, $p < 0.001$). There was also a significant difference between the Christian and the None-or-Other participants ($U=222.5$, $p < 0.05$). It was also found that consumer religio-centrism varies with religious identification in our sample (Chi-Square(2)=51.7, $p < 0.001$). The significant differences are between the Muslim participants and Christian participants ($U=69.5$, $p < 0.001$), and between the Muslim and the None-or-Other religion participants ($U=49.5$, $p < 0.001$).

4.2.3 Procedure and experimental tasks

All experimental tasks during this study were performed in a laboratory at Brunel University to control for any effects due to physical location. Each participant interacted with three Web sites; a same-religion site, a different-religion site and a neutral site. The neutral site in both cases was www.bol.com. The same-religion site for the Muslim participants was www.dar-us-salam.com (the other-religion site for Christian participants); for Christian participants the same-religion site was www.christianbooks.com (the other-religion site for the Muslim participants). They were asked to explore each web site for approximately 20 minutes before filling out a questionnaire based on their experience of it. Following completion of this part of the study, participants were asked to complete the demographic questionnaire.

4.3 Results

4.3.1 Results for web-based trust

The mean ratings of Web-based trust for each site by each group of participants are summarised in Table 3. A 3 x 3 mixed-factorial analysis of covariance was performed on ratings of web-based trust (see Table 4). The within-subjects independent variable consisted of displayed religion of the web-site. The between-subjects independent variable was religious identification of the participant. Gender was used as a covariate to remove effects due to the uneven distribution of gender of the sample. Analyses were performed by SPSS using type III adjustment for unequal sample size. After adjustments by the covariate, a significant multivariate main effect was observed with $F(2, 86) = 3.67, p < 0.05$. There was also a significant interaction effect with religious identification of the participant ($F(4, 86) =$

5.15, $p < 0.05$). A test of the within-subjects effects showed a significant main effect of site religion ($F(1, 87) = 6.75, p < 0.05$) and a significant interaction effect between site religion and religious identification ($F(2, 87) = 6.0, p < 0.05$). There were no significant between-subjects effects. Planned comparisons were conducted for the Muslim and Christian portion of the sample to test Hypotheses H1a and H2b that in-group members of a religion will trust their own religion Web-site more than a neutral or other religion web-site. It was found that the Muslim participants did express significantly more trust in the Muslim book site compared to the Christian book site ($t(37) = -2.74, p < 0.05$), however there was no significant difference between the mean trust rating for the Muslim site and the Neutral site. No significant differences were found in the Christian sub-sample so we cannot reject the null hypothesis.

TABLE 3: Mean ratings of web-based trust.

Participant religious identification	Site Religion:		
	Christian	Muslim	Neutral
Christian	3.49 (s.d. 1.04)	3.21 (s.d. 1.03)	3.49 (s.d. 0.86)
Muslim	3.18 (s.d. 0.88)	3.62 (s.d. 0.78)	3.39 (s.d. 0.80)
None-or-Other	3.44 (s.d. 0.75)	3.06 (s.d. 0.54)	3.60 (s.d. 0.85)

TABLE 4: Analysis of covariance of web-based trust.

Source	Sum of Squares	df	Mean Square	F	Sig.
Site religion	3.677	1	3.677	4.529	$p < 0.05$
Site religion x gender	2.635	1	2.635	3.246	n.s.
Site religion x religious identification	9.152	2	4.576	5.635	$p < 0.05$
Error (site)	70.643	87	0.812		

TABLE 5: Mean ratings of web-based attitude.

Participant religious identification	Site Religion:		
	Christian	Muslim	Neutral
Christian	2.50 (s.d. 0.94)	2.22 (s.d. 0.69)	3.39 (s.d. 0.65)
Muslim	2.39 (s.d. 0.77)	3.85 (s.d. 0.94)	3.18 (s.d. 0.92)
None-or-Other	2.85 (s.d. 0.70)	2.53 (s.d. 0.51)	3.78 (s.d. 0.71)

TABLE 6: Analysis of covariance of web-based attitude.

Source	Sum of Squares	df	Mean Square	F	Sig.
Site religion	8.572	1	8.572	16.906	$p < 0.001$
Site religion x gender	3.112	1	3.112	6.138	n.s.
Site religion x religious identification	42.309	2	21.154	41.721	$p < 0.001$
Error (site)	44.113	87	0.507		

4.3.2 Results for web-based attitude

The mean ratings of Web-based attitude for each site by each group of participants are summarised in Table 5. A 3 x 3 mixed-factorial analysis of covariance was performed on ratings of attitude toward the web sites (see Table 6). The within-subjects independent variable consisted of displayed religion of the web-site. The between-subjects independent variable was religious identification of the participant. The covariate was gender. Analyses were again performed using type III adjustment for unequal sample size. After adjustments by the covariate, a significant multivariate main effect was observed with $F(2, 86) = 10.65, p < 0.001$. There

was also a significant interaction effect with religious identification of the participant ($F(4, 86) = 5.65, p < 0.001$). A test of the within-subjects effects showed a significant main effect of site religion ($F(1, 87) = 13.57, p < 0.001$) and a significant interaction effect between site religion and religious identification ($F(2, 87) = 11.64, p < 0.001$). There were no significant between-subjects effects. Within-subjects planned comparisons were conducted with the Muslim and Christian portion of the sample to test hypothesis H2a and H2b that in-group members of a religion would express more positive attitude towards their own religion site compared to another religion site or a neutral site. It was found that the Muslim participants did express

significantly more positive attitudes towards the Muslim site compared to either the Christian site ($t(37) = -7.752$, $p < 0.001$) or the Neutral site ($t(37) = -2.13$, $p < 0.05$).

For the Christian portion, a significant difference was found between attitude towards the Christian site and the Neutral site, but this was in the opposite direction to that predicted, with the neutral site being preferred over the Christian site ($t(28) = -4.37$, $p < 0.001$). No significant difference was found between attitude towards the Christian and the Muslim sites.

5. DISCUSSION AND CONCLUSIONS

The global nature of the internet allows e-commerce stores to market around the world. Within this context it is important for web retailers to understand how aspects of their sites may help to engender trust and positive attitude within members of particular cultural groups. The current paper has explored the role which consumer cultural identification plays in the initial formation of trust and attitude towards Web-based stores which display culture-specific characteristics.

The main contribution of this paper is to show that trust and attitude in a web-based retailer can vary with culture, at least as represented by affiliation to a religion. Within one collectivist culture, identification with the Muslim religion, we have shown that in-group trust effects can transfer to the web context. Muslims in our study were found to be more trusting of a Muslim site compared to a Christian site (though not compared to the Neutral site). Muslim participants also liked the Muslim site more than the neutral and the Christian site. Since the attitude construct used in this study is closely related to willingness to buy, this finding suggests that this group would be more likely to buy from the Muslim site than the other two sites tested.

However, the hypothesised relationship between trust and religious identification was not supported by the Christian participants. The Christian portion liked the neutral site better than the Christian site. Overall the pattern of responses given by the Christians was similar to those given by the None-or-Other religion group. This could imply that Christians do not use religious distinctions in forming their initial trust and attitude towards a web site. In the absence of culture effects we would expect large and reputable stores (such as Bol.com) to be trusted and liked better by participants [15].

The lack of predicted effects with the Christian sample calls into question the generalisability of the findings with the Muslim sample to other collectivist religions. However, a potential explanation for the different results between the Christian and Muslim samples can be attributed to the scores of these two groups on the religious commitment scale which differed significantly, so may be the Christians in our sample were not

“devout” Christians. Interestingly the Muslim and Christian participants also differed significantly in terms of how much they said religion affected their general purchasing decisions (as measured by the RELSCALE). This supports the idea that the effects we have observed in our experiment may demonstrate a transfer of in-group attitudes from the general to the web context.

5.1 Limitations and future work

There are a number of limitations that were identified during the course of this research study. One prominent issue was that we have assumed the existence of homogeneity within the religious groups. However, both Christianity and Islam have experienced profound schisms that resulted in the inception of several religious factions such as Roman Catholics and Eastern Orthodox in Christianity, and the Sunni and Shia sects in Islam [9]; Future work could address this issue by concentrating on specific religious sects, though the practical applicability of such research might be limited. Other avenues for future research include investigating online religious trust relationships for religious subgroups in (1) individualistic societies (e.g. US) and (2) in situations of high-risk purchase decisions (e.g. books) versus low-risk purchase decisions (e.g. cars) [20].

REFERENCES

- [1] Bearden, W.O. and Netemeyer, R.G., *Handbook of Marketing Scales* (Second edn.). Thousand Oaks: Sage Publications, 1999.
- [2] Brewer, M.B., In-group bias in the minimal intergroup situation: a cognitive-motivational analysis. *Psychology Bulletin*, Vol. 86, pp307-324, 1979.
- [3] Butler, J.K., Towards Understanding and Measuring Conditions of Trust: Evolution of a Conditions of Trust Inventory. *Journal of Management*, Vol. 17, pp643-663, 1991.
- [4] Deutsch, M., Trust and suspicion. *Journal of Conflict Resolution*, Vol. 2, pp265-279, 1958.
- [5] DeVellis, R.F., *Scale Development: Theory and Applications*. London: Sage Publications, 1991.
- [6] Egger, F.N., Trust me, I'm an online vendor: towards a model of trust for e-commerce system design. *CHI 2002*, 2002.
- [7] Giffin, K., The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. *Psychological Bulletin*, Vol. 68, pp104-120, 1967.
- [8] Hoffman, D.L., Novak, T.P., and Peralta, M., Building consumer trust online. *Communications of the ACM*, Vol. 42, pp80-85, 1999.
- [9] Hofstede, G., *Cultures and Organizations*. London: Harper Collins Business, 1998.
- [10] Hosmer, L.T., Trust: the connection link between organizational theory and philosophical

- ethics. *Academy of Management Review*, Vol 20, pp379-403, 1995.
- [11] Howard, J.A., *Buyer Behavior in Marketing Strategy* (Second ed.). London: Prentice-Hall, 1994.
- [12] Husted, B.W., Trust in business relations: directions for empirical research. *Business and Professional Ethics Journal*, Vol. 8, pp23-40, 1989.
- [13] Iannaccone, L.R., Risk, rationality and religious portfolios. *Economic Inquiry*, Vol. 13, pp285-295, 1995.
- [14] Igarria, M. and Parasuraman, S., Attitudes towards microcomputers: development and construct validation of a measure. *International Journal of Man-Machine Studies*, Vol. 35, pp553-573, 1991.
- [15] Jarvenpaa, S. L. and Tractinsky, N., Consumer trust in an internet store: a cross-cultural validation. *Journal of Computer-Mediated Communication*, Vol. 5, No. 2, 1999.
- [16] Jarvenpaa, S.L., Tractinsky, N. and Vitale, M., Consumer trust in an internet store. *Information Technology and Management*, Vol. 1, No. 1-2, pp45-71, 2000.
- [17] Jones, G.R. and George, J.M., The experience and evolution of trust: implications for cooperation and teamwork. *Academy of Management Review*, Vol. 23, No. 3, pp531-547, 1998.
- [18] Kim, J. and Moon, J.Y., Designing towards emotional usability in customer interfaces -- trustworthiness of cyberbanking system interfaces. *Interacting with Computers*, Vol. 10, pp1-29, 1998.
- [19] Lee, M. and Turban, E., A trust model for consumer Internet shopping. *International Journal of Electronic Commerce*, Vol. 6, No. 1, pp75-92, 2000.
- [20] Lowengart, O. and Tractinsky, N., Differential effects of product category on shoppers' selection of web-based stores: a probabilistic modeling approach. *Journal of Electronic Commerce Research*, Vol. 2, No. 4, pp12-26, 2001.
- [21] Michalos, A., The impact of trust on business, international security, and the quality of life. *Journal of Business Ethics*, Vol. 9, pp619-638, 1990.
- [22] O'Keefe, R.M., Cole, M., Chau, P.Y.K., Massey, A., Montoya-Weiss, M., and Perry, M., From the user interface to the consumer interface: results from a global experiment. *International Journal of Human-Computer Studies*, Vol. 53, pp611-628, 2000.
- [23] Pearce, W.B., Trust in Interpersonal Relationships. *Speech Monographs*, Vol. 41, pp236-244, 1947.
- [24] Proctor, T., *Essentials of Marketing Research*. London. Pitman Publishing, Pearson Professional Ltd, 1997.
- [25] Rotter, J.B., A new scale for the measurement of trust. *Journal of Personality*, Vol. 35, pp651-665, 1967.
- [26] Shimp, T.A. and Sharma, S., Consumer ethnocentrism: construction and validation of the CETSCALE. *Journal of Marketing Research*, Vol. 24, pp280-289, 1987.
- [27] Solomon, M.R., *Consumer Behavior: Buying, Having, and Being* (Fourth edn.). Upper Saddle River, New Jersey: Prentice Hall, Inc., 1999.
- [28] Stark, Rwanda Glock, C.Y., *American Piety: The Nature of Religious Commitment*. California: University of California Press, 1968.
- [29] Staw, B.M. and Cummings, L.L., *Research in Organizational Behavior: An Annual Series of Analytical Essays and Critical Reviews*. Greenwich, CT: JAI Press, 1980.
- [30] Tabachnick, B.G. and Fidell, L.S., *Using Multivariate Statistics* (Third edn.). NY: HarperCollins College Publishers, 1996.
- [31] Tan, Y-H. and Thoen, W., Towards a generic model of trust for electronic commerce. *International Journal of Electronic Commerce*, Vol. 5, No. 2, pp61-74, 2000.
- [32] Triandis, H.C., *Individualism and Collectivism*. Boulder, Colorado: Westview Press, 1995.
- [33] Turner, J.C., *Rediscovering the Social Group: A Self-categorization Theory*. Oxford: Blackwell, 1987.
- [34] Wang, H., Lee, M.K.O., and Wang, C., Consumer privacy concerns about internet marketing. *Communications of the ACM*, Vol. 41, pp63-70, 1998.
- [35] Zikmund, W.G., *Business Research Methods* (Fifth edn.). Orlando: The Dryden Press, Harcourt Brace College Publishers, 1997.
- [36] Zeynep, G.-C. and Durairaj, M., Cultural variations in country of origin effects. *Journal of Marketing Research*, Vol. 37, pp309-327, 2000.

Trust for E-Business Management

Liaquat Hossain¹, Rolf T. Wigand²

¹School of Information Technologies, G82C Madsen Building F09,
The University of Sydney, NSW 2006 Australia

²Information Science and Management, Department of Information Science, CyberCollege
University of Arkansas at Little Rock, Little Rock, AR 72204-1099, USA
lhossain@it.usyd.edu.au, rtwigand@ualr.edu

ABSTRACT

How do we develop and sustain trust? What is the process for building trust between business partners in virtual environments? Is there a significant difference between the development and sustainability of trust online or offline? In this paper, we first introduce the concept of e-business and discuss the importance of trust for ensuring effective collaboration. Secondly, we discuss the relationships between e-collaboration and trust for managing e-business. Thirdly, we suggest a framework, which may help facilitate the development and sustainability of trust in an online environment. Finally, implications for the development and sustainability of trust online, which can be used to understand the interplay among technologies, e-business and collaboration is provided. We suggest that the implications of this study are three-fold: trustworthy relationships among business partners, effective sustainable collaboration, and optimal use of ICT for supporting e-business activities.

Keywords: Trust, Social exchange, Collaboration, e-Business

1. INTRODUCTION

In general, e-business refer to the use of Internet technologies for support activities such as buying, selling products and services, as well as collaboration with business partners. We first highlight that e-business is not a synonym for e-commerce. E-business refers to an organization's capability to electronically connect with their partners for exchanging values such as knowledge, products and services (Venkatraman and Henderson, 1998; Hackbarth and Kettinger, 2000; Fahey, et al., 2001; Flurry and Vicknair, 2001; and Davis and Ritsko, 2001). Exchanging values with business partners requires collaboration. Therefore, collaboration with business partners is increasingly seen as a critical aspect of e-business management (Davis and Ritsko, 2001; Sairamesh et al 2002; Lechner and Hummel, 2002; Alt and Fleisch, 2001). Hackbarth and Kettinger (2000) further highlight that e-business should address issues such as human performance within the organization and also in an inter-organizational network. Therefore, understanding the exchange relationships among buyer, suppliers and trading partners is seen central for e-business management (Markus et al 2000). We argue that the management of e-business requires understanding strategies for building effective collaboration, and ensuring the optimal use of Internet infrastructure for supporting the sharing of knowledge among buyers, suppliers and trading partners. We suggest that for an e-business to be successful, an effective collaboration must exist among different participating business partners. Therefore, the design of the e-business systems should be based on a framework, which facilitates the development and sustainability of collaboration among business partners.

We emphasize that collaboration is essentially a social phenomenon and is regarded as a complex issue. IT is there to support the flow of effective sharing only when the partners have adequate level of trust so that they can start to collaborate with each other. We suggest that a low level of trust may lead to a lower level of IT utilization for supporting collaboration and higher level of trust between business partners would lead to a higher level of collaboration so that the use and utility of IT can be maximized. So, why is collaboration a complex social phenomenon? How do we build an effective collaborative culture so that the use and utility of IT can be maximized? Here, we present a casual model, which can serve as a basis for the management of e-business activities and also to ensure the utilization of ICT for supporting the relationships among business partners. Discussions of the interrelationships of different variables and their dependency are presented in the forthcoming section.

The focus of this study is on understanding the relationships among the use of Internet technologies, the degree of collaboration and its social context for the management of e-business activities. In the next section, we suggest the importance of collaboration in managing e-business activities. We further discuss the complexity of collaboration and in particular, e-collaboration for managing e-business activities. We apply social exchange theory to develop an understanding of the phenomenon referred as social exchange on-line. Here, trust is considered to be the key antecedent and discussed in relation to supporting effective collaboration. We conclude that social exchange theory can be applied to further our understanding of how structure can be used in facilitating the development and sustainability of trust.

Finally, a framework for the development and sustainability of trust on-line is provided.

2. SUPPORTING COLLABORATION THROUGH ICT

Collaboration, in its traditional sense, is the collective interaction between multiple parties. Rapid advances in technology along with globalization provides opportunities and challenges for businesses to establish collaborative relationships with suppliers and customers so that they can enhance competitiveness and lower costs (Karsten, 1999). In e-business terms, collaboration connects all the parties involved in the process, which enables information-sharing. This facilitates structured and controlled interactions through proper channel-connectivity, electronic information and workflow. Connectivity provides the core network for interaction. These network systems not only eliminate regional, inter-departmental and intra-enterprise barriers, but they also bring together a network of people around specific processes. Collaboration tools such as the TeamSCOPE system trail is proven to be effective Team Software for a geographically dispersed collaborative project environment (Steinfeld et al., 1999). TeamSCOPE, a web-based trail system provides each team with a team-shared file repository, which makes it easy to users to store and exchange group-related files.

Once connectivity is established, the next major aspect is sharing information because electronic information acts as the lifeblood of collaboration in organization. For example, previous studies suggest that the collaborative culture is an essential prerequisite successful use of IT for supporting sharing (Orlikowski, 1993; Bowers, 1994; and Gallivan et al., 1993). We suggest that three core elements--connectivity, electronic information and workflow make e-business collaboration a very powerful way of bringing a group of people together to interact and communicate efficiently. In this regard, Vandenbosch and Ginzberg (1997) further suggest that organizations must first meet four basic conditions in order for the implementation of IT to support effective collaboration. The first is the organizational members need to collaborate. The second is the understanding of IT and its use in supporting collaboration. The third is the organizational support for the adoption, implementation, and continued use of the technology. The fourth is the organizational culture that supports collaboration.

The purpose of collaborative workspace is to provide a structure or framework, which allows its user to achieve and understand its purpose (Chin et al., 2002). In addition, it should support the users to adapt and follow conventions for its use. The behavior of a collaborative workspace must be controllable and support the orientation of the members in the workspace and in the cooperation process. We suggest that the development of collaborative workspace requires a high level of attention

to the basic principles of organizational design such as— (i) the connection between the design of e-business and the elements of the social world leads to effective collaboration; (ii) components must be designed in ways that work against the strong tendency towards entropy; and (iii) open access leads to higher collaboration. This can be supported by the case of combustion researchers collaborating over the Internet (see Pancerella et al., 1999) where effective collaboration was achieved through the application of principles of organizational design presented above. We suggest that the effective use of IT for supporting collaboration requires frameworks/models as a guiding principle so that individuals and groups in an organization or inter-organizational network can work effectively and efficiently. We discuss this further by focusing on two popular collaboration models—the *contract* model (Wachter and Reuter, 1992) and the *consortium* model (Kouramajian et al., 1995) and suggest a rich framework called *social exchange* perspective to further understand the social aspect related to effective collaboration for e-business activities.

Collaboration for information sharing and knowledge management is increasingly becoming possible through the introduction of network computing (Kouramajian et al. 1995). Kouramajian et al. (1995) further suggest that the *consortium* provides a flexible framework for defining collaborative transactions. Contrary to this, a *contract* model is designed to provide a programming environment for long-lived transactions between parties concerned (Wachter and Reuter, 1992). Studies suggest that interaction under the *contract* model is rather tightly constrained tightly to collaborative tools and may severely restrict collaboration among users (Wachter and Reuter, 1992; Kouramajian et al., 1995). The *consortium*, on the other hand, consists of both the structural and the behavioral models and provides a new way for collaborative interaction (for example, What You See Is What You Want-WYSIWYW). The structural model is based on the assumption that a large collaborative task can involve multiple phases or sessions and different phases may have different participants. In addition, the behavioral model defines mechanisms for specific collaborative actions. Therefore, two main abstractions—sessions and transactions are integral parts of the consortium framework for supporting e-collaboration. Here, a session provides a convenient abstraction for a collective action. Further, an ordered set of sessions where the owner specifies the order is regarded as part of transactions.

However, these models do not provide an adequate background for understanding the implications of trust as a social factor for the effective utilization of IT in an e-collaborative environment. For example, work by Malone et al. (1987) and Clemons et al. (1993) focused primarily on IT as a means for reducing inter-organizational transaction cost, without taking social factors such as trust and its implications in reducing

transaction costs in inter-organizational systems (Kumar et al., 1998). In the information systems literature, the development and sustainability of trust for managing virtual collaborative relationship is increasingly seen important for ensuring the optimal use of IT for supporting e-business activities (Whittaker, 1996; Hoffman et al., 1999; Jones et al., 2000; Ratnasingham and Kumar, 2000; Ono et al., 2001; Nayak et al., 2001; Kasper-Fuehrer et al., 2001). Recent studies on the use of computer-mediated communication tools for supporting trust building in an on-line environment suggest that audio and video technologies have the potential to be as effective as face-to-face communication as long as the design of systems is guided by the social process in place (Bos et al., 2002). In line with these we suggest the following propositions:

Proposition 1:

Higher collaboration relates to optimal use of ICT for supporting e-business.

Proposition 2:

There is no direct relationship between the level of collaboration and level of ICT use for supporting e-business activities.

Proposition 3:

Sessions and transactions are directly related to higher level of collaboration.

3. UNDERSTANDING THE RELATIONSHIPS BETWEEN TRUST AND COLLABORATION

Hosmer (1995) suggests that trust deals with the expectation by one person, group, or firm of ethical behavior which is morally right decisions and actions based upon ethical principles of analysis on the part of the other person or part in an exchange. Therefore, ensuring successful business transactions and cooperative work requires a high level of attention to the development and sustainability of trust (Josang, 1996; Nelson and Coopriider, 1996; Sheppard and Tuchinsky, 1996; Swagerman et al., 2000; Wigand, 1978). Trust can also be based upon the rational appraisal of a partner's reliability and competence, and upon feelings of concern and attraction. These factors may help to reduce awkwardness, complexity and uncertainty in social interactions and therefore, make the collaboration effective.

A study by Ratnasingham and Kumar (2000) suggest that trust and interpersonal relationships are essential elements for understanding the role of communication media in collaborative work. Collaboration is seen as most effective and rewarding when the participants trust each other (Axelrod, 1997). We suggest that two types of activities--*cognitive and emotional* faculties must be established for developing trust. Cognitive-oriented activities may convey competence and reliability, and thereby increase confidence that a task will be

successfully completed. Emotion-oriented activities can create an emotional bond, and help decrease fears of exploitation and increase feelings of mutual support for building trust.

Recent studies on interorganizational networks, virtual organizations and the long-term sustainability of virtual collaboration suggest that trust is critical so that the IT infrastructure can be leveraged optimally to support the exchange among business partners (Josang, 1996; Whittaker, 1996; Bandow, 1998; Staples and Ratnasingham, 1998; Karahannas and Jones, 1999; Ratnasingham and Kumar, 2000; McKinght et al. 2000; Kasper-Fuehrer et al., 2001; Nayak et al., 2001; Ono et al., 2001; Bos et al., 2002; Gefen, 2002). We argue here that the development and sustainability of trust in an on-line or virtual environment requires an understanding of the social exchange systems in which individuals, groups or organizations operate. This is important as effective collaboration is seen to be largely dependent on the trust relations among different interest groups (Abdul-Rahman and Hailes, 1997).

It is therefore suggested that there is a direct relationship between the success of e-business and the embedded social exchange systems under which individuals, groups or organizations operate. We argue that this social exchange system helps building a collaborative and sharing culture in organizations. Chadwick-Jones (1976) also suggests that the level of social exchange is dependent on trust. Blau (1964) highlights that distrust is expected in economic relations but have a negative impact on social behavior. Thibaut and Kelley (1959) and Blau (1964) further suggest that building trust is a gradual process, which requires a cumulative commitment to a relationship. For example, the Prisoner's Dilemma game illustrates the problems associated with building trust. In this game, the participants are given choices between only two alternatives—(i) sharing the outcomes equally with another person or (ii) try to maximize his/her own gains at the expense of the other. On the basis of the aforementioned discussion, we suggest the following propositions related to collaboration and trust:

Proposition 4:

Higher levels of trust lead to higher collaboration among business partners.

Proposition 5:

Lack of trust or, conversely, higher levels of distrust lead to a lower level of collaboration among business partners.

Proposition 6:

High competency and reliability of business partners lead to high degree of trust.

Proposition 7:

There is no direct relationship between the level of competency and reliability of business partner for developing the level of trust.

Therefore, common business goals or a shared vision is considered to be a prerequisite for ensuring the optimal use of IT systems for supporting information exchange and knowledge sharing among business partners in an inter-organizational network (Wigand et al., 1997; Wigand, 1977; Scott and Lane, 2000). We suggest that the social exchange perspective presented below provides a useful mechanism for understanding the complex process of building trust such that higher collaboration among business partners may be achieved. Moreover, embedded in the social exchange perspective is a social control mechanisms that mutually regulates exploitive and opportunistic behavior (such as suggested in the Prisoner’s Dilemma discussion above). This may in fact reduce uncertainty by guarding against opportunistic behavior among participating members (Fichman, 1997; Wicks et al., 1999; Grambowski and Roberts, 1999).

It follows then to ask, how do we build and sustain trust through social exchange? It is evident from the previous studies on ICT, trust and collaboration that understanding social exchange systems is a powerful mechanism for developing an understanding of the development and sustainability of trust among the business partners. Bandow (1998) suggests the importance for exploring strategies for developing trust in virtual environment, which is now a reality due to increasing use of IT in the work place. This deserves careful attention as IT may be used to allow or support work in separate locations, the relationships with others in their groups are not the same due to the reduced or non existent of face-to-face contact. A recent empirical study of the effectiveness of computer mediated communication channels of trust development suggest that video and audio conferencing groups are nearly as good as face-to-face (Bos et al., 2002). However, the use of IT such as video and audio for building trust is considered to be a slower progress towards a full cooperation and vulnerable to opportunistic behavior (Das and Teng, 1998).

4. TRUST THROUGH SOCIAL EXCHANGE

Rapoport and Guyer (1966) suggest seventy-eight forms of two-person games where the mixed-motive can be the impediments for building trust. This Prisoner’s Dilemma design originated from the anecdote about the two persons arrested on suspicion of committing a crime (Chadwick-Jones, 1976). Here, two suspects are taken into custody and separated with an assumption that they are guilty of a specific crime (Luce and Raiffa, 1957). However, the district attorney does not have enough evidence against the suspects so that they can be convicted at a trial. Therefore, the district attorney suggests two alternatives to each prisoner—(i) to confess to the crime or (ii) not to confess. Table 1 provides an example of the Prisoner’s Dilemma game. It is suggested in the dilemma matrix presented below that the first entry in each cell below represents the choice (and payoff) for

Prisoner A and the second entry represents the outcome for Prisoner B.

		Prisoner B	
Prisoner A		Not Confess	Confess
Not Confess		C, C 1 year, 1 year (Cooperate, Cooperate)	C, D 9 years, 2 months (Sucker, Temptation)
Confess		D, C 2 months, 9 years (Temptation, Sucker)	D, D 6 years, 6 years (Defect, Defect)

Table 1. An example of the Prisoner’s Dilemma game (Chadwick-Jones, 1976)

Therefore, both A and B receive one year of imprisonment in the cooperation cell and in the joint defection option both receive several years. However, in the temptation cell, Prisoner A gets off with only two months, whereas Prisoner B refusing to confess, receives the worst type of sentence (the sucker option). In this game, the player A and B may decide to cooperate or may decide to go for ‘temptation cell’ for gaining more for one. An intention to gain more for one may lead to ‘defecting’ from the cooperative option to a competitive one. The danger behind moving from cooperation to competition is the loss of trust, which may lead to distrust between A or B. We suggest that the higher the level of distrust between individuals or parties, the lesser the level of collaboration. It is also suggested in the social psychology literature that the higher-level distrust may lead to defensive behavior and both parties may lose sustainable collaboration (Chadwick-Jones, 1976).

We suggest that understanding the dynamics of reciprocity and imbalance is important for the trust building process, which in fact would ensure higher levels of collaboration. Blau (1964) was interested in understanding the complexity of attracting individuals to a relationship and suggested that one may be attracted by intrinsic or by extrinsic benefits. Table 2 below provides an illustration of Blau’s (1964) view of relationship formation. Here, a distinction between reciprocal and unilateral transactions is drawn. Blau (1964) further argued that the distinction between reciprocal and unilateral transactions suggest a dynamic force, which transforms a social process simple into an increasingly complex one.

For example, intrinsic rewards such as love in a reciprocal relationship may suggest mutual attraction and the extrinsic rewards such as advice in carrying out a task or a reciprocal relationship may suggest exchange between individuals or parties involved. Blau (1964) further highlights that the exchange of extrinsic benefits helps sustain the major intrinsic rewards when an association is intrinsically rewarding. Similarly, one-sided attachment is established when intrinsic benefits are part of unilateral relationships. However, in the case

of unilateral extrinsic benefits, power is established because those who cannot reciprocate in kind are obligated to comply with the other person's wishes (Chadwick-Jones, 1976). In line with this, we suggest the following proposition:

Proposition 8:

Trust building is dependent on the dynamics of reciprocity and imbalance between business partners.

	Intrinsic	Extrinsic
Reciprocal	Mutual attraction	Exchange
Unilateral	One-sided attachment	Power

Table 2. Illustration of a two-dimensional distinction among transactions

From another point of view, norms or agreements are important in trust-building processes as they help avoid uncertainty (Chadwick-Jones, 1976). For example, the study of kingship relations suggest that interpersonal trust reduces one principle source of uncertainty in actual social situations (Anderson, 1971). Anderson (1971) argues that uncertainty in social relationships influences someone to be calculative and also to seek shorter-term returns. The study of kingship relation suggested by Anderson (1971) is based primarily on two assumptions—(i) calculative behavior leads to increased variability in behavior within business partners under the extreme uncertainty in social or kingship relations; and (ii) arrival of calculative orientations in a community may lead to termination of exchange which requires high levels of trust. It is further suggested that the general norms of kingship behavior will be a reality only where communities are highly homogeneous and stable (Anderson, 1971).

Therefore, it may be concluded that social exchange theory is based on the assumption that in order for exchanges of goods and behaviors to take place in an organization, there is a joint activity or transaction, which would involve two or more actors when each of them has something the other would value (Thibault and Kelley, 1959; Homans, 1961; Emerson, 1972). Here, the assumption made is that actors face substantial degrees of uncertainty and ambiguity about what the potential partner value, the utility of different exchanges to them and what exchanges are been made among others in the exchange network (Molm and Cook, 1995). Social exchange theory primarily deals with how actors react to these uncertainties and ambiguities, in which the actors based their expectations of rewards, cost and punishments. Likewise, the actors also use this information to anticipate their future exchanges in terms of the rewards, cost and punishments.

In addition, this theory highlights the importance of cost and rewards in the formation of social relationship. According to Thibault and Kelley (1959), exchange theory is based on the exchange of rewards and costs to quantify the values of outcomes from different situations for an individual. People strive to minimize costs and maximize rewards and then base the likeliness of developing a relationship with someone on the perceived possible outcomes. For example, when a person perceives a greater outcome from another person, he or she intends to disclose or share the knowledge greater and thus develops a closer relationship with that person. Another important aspect of this theory is that people will establish a reciprocal relationship among themselves once these two main elements—cost and rewards are fulfilled. We discuss the implications of this study in the next section.

5. IMPLICATIONS FOR MANAGING E-BUSINESS

The implications of this study are three-fold—trustworthy relationships among business partners, effective sustainable collaboration, and optimal use of ICT for supporting e-business activities.

5.1 Trustworthy relationships among business partners

Here, trustworthy relationships refer to non-symmetrical or unidirectional and conditionally transitive agreements between parties. Trust relations may be divided into two categories (Abdul-Rahman and Hailes, 1997)—*direct trust relationship* and *recommender trust relationship*. For example, there is a *direct trust relation* if person A trusts person B. However, if person B trusts person A to give recommendations about other persons' trustworthiness, then there is a *recommender trust relationship* between person A and person B. For the case of virtual networks, the opportunities for face-to-face interactions to facilitate this is minimal. Therefore, the virtual environment constitutes a special case of network organizations and is characterized by lateral rather than vertical relationships (Snow et al., 1992; Bleecker, 1994; Semich, 1994; Garreth, 1998 and Wigand et al., 1997). Kasper-Fuehrer et al., (2001) suggests that authority is clearly defined by the hierarchical structure in the organizational design, along with organization charts and formalized procedures in vertical relationships. However, this hierarchical coordination and control mechanisms is lacking in a network organization. Therefore, trust may act as a substitute for control in a network or lateral relationship organizations (Sheppard and Tuchinsky, 1996; Handy, 1995; Jones and Bowie, 1998).

5.2 Effective Sustainable Collaboration

The authors emphasize that the arguments for a social exchange perspective presented in the previous section provides valuable insights about the trust building

process for the sustainability of virtual network relationships. We argue that this perspective of e-business management is important in order for exchanges of goods and behaviors to take place in an organization. This would also encourage joint activity or transaction, which would involve two or more actors when each of them has something the other would value (Thibault and Kelley, 1959; Homans, 1961; Emerson, 1972). However, it is important to note that actors face substantial degrees of uncertainty and ambiguity about what the potential partner value, the utility of different exchanges to them and what exchanges are being made between others in the exchange network (Molm and Cook, 1995). We conclude that social exchange theory (Chadwick-Jones, 1976) is a powerful mechanism for understanding the relationship building process as it primarily deals with how actors react to these uncertainties and ambiguities, in which the actors based their expectations of rewards, costs and punishments and also to anticipate their future exchanges.

5.3 ICT Use for Supporting e-Business

How may a higher level of trust help sustain higher levels of collaboration among business partners? We suggest that higher levels of trust among parties may be achieved by the norms or agreements as they help reduce uncertainty (Anderson, 1971; Chadwick-Jones, 1976) and help establish common business goals or shared business vision among parties involved (Wigand et al., 1997; Scott and Lane, 2000). Studies suggest that this shared vision helps in understanding each member's role, identification of groups, determination of critical behaviors such as willingness to cooperate with others, and also willingness to engage in mutual goal setting, so that higher levels of collaboration may be achieved among partners (Albert and Whetton, 1985; Wiesenfeld et al., 1998). Therefore, establishing group identity, awareness of mutual needs and expectations, and clarification of tasks and responsibilities is a necessary prerequisite for building trust and also to promote collaboration (Anderson, 1971; Kasper-Fuehrer et al. 2001). We further suggest that the use and utility of ICT may be optimal for supporting e-business activities once these necessary prerequisites are used to promote collaborative culture during the design phase of business systems (Gallivan et al., 1993; Orlikowski, 1993; Bowers, 1994).

6. CONCLUSION

We utilized social exchange theory to explore the underlying relationships among ICT, trust, and collaboration as a means to advance our understanding of the management of e-business activities. We have argued three important factors so that ICT may be effectively used to support collaboration and suggested sets of propositions. Firstly, we suggest that the successful integration of this architecture for a network organization is largely dependent on the establishment of a shared

business vision. This would help build trust among business partners so that a higher level of e-collaboration may be achieved. Secondly, we suggest that social factors such as trust are a key element for ensuring the optimal use of ICT so that higher levels of collaboration can be achieved among business partners. Thirdly, we introduce the game-theoretic approach of the *Prisoner's Dilemma* to argue that distrust among business partners is the result of the adverse consequences of calculative and opportunistic behavior. We conclude that this distrust among business partners may lead to the termination of exchanges and impede the possibility of e-collaboration.

REFERENCES

- [1] Abdul-Rahman, A., and Hailes, S. (1997). 'A distributed trust model,' *Proceedings of the 1997 Workshop on New Security Paradigms*, Langdale, Cumbria, UK, pp. 48-60.
- [2] Albert, S., and Whetton, D. (1985). 'Organizational identity,' In L. L. Cummings and B. M. Staw (Eds.), *Research in Organizational Behavior*, Vol. 7, Greenwich. CT: JAI Press, pp. 263-295.
- [3] Alt, R., and Fleisch, E. (2001). 'Business networking systems: characteristics and lessons learned,' *International Journal of Electronic Commerce*, Vol. 5, No. 2, pp. 7-27.
- [4] Anderson, M. (1971). *Family structure in the 19th century Lancashire*, Cambridge University Press, Cambridge, UK.
- [5] Axelrod, R. (1997). *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*, Princeton University Press, Princeton, New Jersey 08540.
- [6] Bandow, D. (1998). 'Working with the Borg: trust, systems development and dispersed work groups,' *Proceedings of the Conference on Computer Personnel Research*, Boston, Massachusetts, pp. 163-169.
- [7] Blau, P. M. (1964). *Exchange and Power in Social Life*, Wiley and Sons, New York.
- [8] Bleecker, S. E. (1994). 'The virtual organization,' *The Futurist*, Vol. 28, No. 2, pp. 9-14.
- [9] Bos, N., Olson, J., Gergle, D., Olson, G., and Wright, Z. (2002). 'Effects of four computer-mediated communications channels on trust development,' *Proceedings of the SIGCHI Conference on Human Factors in computing systems: changing our world, changing ourselves*, Minneapolis, Minnesota, pp. 135-140.
- [10] Bowers, J. (1994). 'The work to make a network work: studying CSCW in action,' in *Proceedings of CSCW'94*, Chapel Hill, NC, New York: ACM, pp. 287-298.
- [11] Chadwick-Jones, J. K. (1976). *Social Exchange Theory: its structure and influence in social psychology*, Academic Press Inc., New York.
- [12] Chin, G., Myers, J., and Hoyt, D. (2002). 'Social networks in the virtual science laboratory,'

- Communications of the ACM*, August, Vol. 45, No. 8, pp. 87-92.
- [13] Clemons, E. K., Reddi, S. P., and Row, M. C. (1993). 'The impact of information technology on the organization of economic activity: the move of the middle hypothesis,' *Journal of Management Information Systems*, Vol. 10, No. 2, Fall, pp. 9-35.
- [14] Das, T. K., and Teng, B. S. (1998). 'Between trust and control: developing confidence in partner cooperation in alliances,' *Academy of Management Review*, Vol. 23, No. 3, pp. 491-512.
- [15] Davis, A. G., and Ritsko, J. J. (2001). 'Preface,' *IBM Systems Journal*, Vol. 40, No. 1, pp. 2-3.
- [16] Fahey, L., Srivastava, R., and Sharon, J. S. (2001). 'Linking e-business and operating processes: the role of knowledge management,' *IBM Systems Journal*, Vol. 40, No. 4, pp. 889-907.
- [17] Fichman, M. (1997). 'A multilevel analysis of trust in inter-organizational customer-supplier ties,' *Proceedings of the Annual Academy of Management Meeting*, Boston, MA.
- [18] Emerson, R. (1972). Exchange theory, part 1: A psychological basis for social exchange, in J. Berger, M. Zelditch, and B. Anderson (Eds.), *Sociological theories in progress*, 2, Boston: Houghton-Muffin.
- [19] Flurry, G., and Vicknair, W. (2001). 'The IBM application framework for e-business,' *IBM Systems Journal*, Vol. 40, No. 1, pp. 8-24.
- [20] Gallivan, M., Goh, G. H., Hitt, L. M., and Wyner, G. (1993). *Incident tracking at infocorp: case study of a pilot NOTES implementation*, MIT Center for Coordination Science, Working paper No. 149, Cambridge, MA: MIT Sloan School.
- [21] Garreth, M. (1998). 'The origin of virtual corporations,' In H. Nolte (Ed.), *Aspects of resource-oriented corporate governance*, Mering, Germany: Rainer Hampp Verlag, pp. 110-133.
- [22] Gefen, D. (2002). 'Reflections on the dimensions of trust and trustworthiness among online consumers,' *ACM SIGMIS Database*, Vol. 33, No. 3, pp. 38-53.
- [23] Grambowski, M., and Roberts, K. H. (1999). 'Risk mitigation in virtual organizations,' *Organization Science*, Vol. 10, pp. 704-721.
- [24] Hackbarth, G., and Kettinger, W. J. (2000). 'Building an e-business strategy,' *Information Systems Management*, Vol. 17, No. 3, pp. 78-93.
- [25] Handy, C. (1995). 'Trust and virtual organization: how do we manage people whom you do not see,' *Harvard Business Review*, Vol. 73, No. 3, pp. 40-50.
- [26] Hoffman, D. L., Novak, T. P., and Peralta, M. (1999). 'Building consumer trust on-line,' *Communications of the ACM*, Vol. 42, No. 4, April, pp. 80-85.
- [27] Homans, G.L. (1961). *Social behavior: Its elementary forms*. New York: Harcourt Brace
- [28] Jovanovich.
- [29] Hosmer, L. (1995). 'Trust: the connection link between organizational theory and philosophical ethics,' *Academy of Management Review*, Vol. 20, pp. 379-403.
- [30] Jones, T. M., and Bowie, N. E. (1998). 'Moral hazards on the road to the "virtual" corporation,' *Business Ethics Quarterly*, Vol. 8, pp. 273-292.
- [31] Jones, S., Wilikens, M., Morris, P., and Masera, M. (2000). 'Trust requirements in e-business,' *Communications of the ACM*, Vol. 43, No. 12, pp. 81-87.
- [32] Josang, A. (1996). 'The right type of trust for distributed systems,' *Proceedings of the ACM 1996 Workshop on New Security Paradigms*, Lake Arrowhead, California, pp. 119-131.
- [33] Karahannas, M. V., and Jones, M. (1999). 'Interorganizational systems and trust in strategic alliance,' *Proceedings of the 20th International Conference on Information Systems*, Charlotte, North Carolina, pp. 346-357.
- [34] Karsten, H. (1999). 'Collaboration and collaborative information technologies: a review of the evidence,' *ACM SIGMIS Database Special Issue on Information Systems: current issue and future changes*, ACM Press, NY, pp. 44-65.
- [35] Kasper-Fuehrer, E. C., Ashkanasy, N. M., and Neal, M. (2001). 'Communicating trustworthiness and building trust in interorganizational virtual organizations,' *Journal of Management*, Vol. 27, No. 3, pp. 235-254.
- [36] Kouramajian, V., Dargahi, R., Fowler, J., and Baker, D. (1995). 'Consortium: a framework for transactions in collaborative environments,' *Proceedings of the 4th International Conference on Information and Knowledge Management*, Baltimore, Maryland, pp. 260-265.
- [37] Kumar, K., van Dissel, H. G., and Bielli, P. (1998). 'The merchant of prato-revisited: toward a third rationality of information systems,' *MIS Quarterly*, Vol. 22, No. 2, June, pp. 199-226.
- [38] Lechner, U., and Hummel, J. (2002). 'Business models and system architectures of virtual communities: from a sociological phenomenon to peer-to-peer architectures,' *International Journal of Electronic Commerce*, Vol. 6, No. 3, pp. 41-53.
- [39] Luce, R. D., and Raiffa, H. (1957). *Games and Decisions*, Wiley and Sons, New York.
- [40] Malone, T. W., Yates, J., and Benjamin, R. I. (1987). 'Electronic markets and electronic hierarchies,' *Communications of the ACM*, Vol. 36, No. 6, pp. 484-497.
- [41] Markus, M. L., Manville, B., and Agres, C. E. (2000). 'What makes a virtual organization work?' *Sloan Management Review*, Vol. 42, No. 1, pp. 13-26.
- [42] McKnight, H. D., Choudhury, V., and Kacmar, C. (2000). 'Trust in e-commerce vendors: a two-stage model,' *Proceedings of the 21st International Conference on Information Systems*, Brisbane, Queensland, pp. 532-536.
- [43] Menon, N. M., Konana, P., Browne, G. J., and Balasubramanian, B. (1999). 'Understanding trustworthiness beliefs in electronic brokerage usage,'

Proceedings of the 20th International Conference on Information Systems, Charlotte, North Carolina, pp. 552-555.

[45] Molm, L., & Cook, K. (1995). Social exchange and exchange networks, in K.S. Cook,

[46] G.A. Fine, and J.S. House (Eds.), *Sociological perspectives on social psychology*, p.209-

[47] 235, Boston: Allyn & Bacon.

[48] Nayak, N., Bhaskaran, K., and Das, R. (2001). 'Virtual enterprises: building blocks for dynamic e-business,' *Proceedings of the workshop on Information Technology for virtual enterprises*, Queensland, Australia, pp. 80-87.

[49] Nelson, K. M, and Coopridge, J. G. (1996). 'The contribution of shared knowledge to IS group performance,' *MIS Quarterly*, Vol. 51, pp. 123-140.

[50] Ono, C., Paulson, B. C., Kanetomo, D., Cutkosky, M., Kim, K., and Petrie, C. J. (2001). 'Trust-based facilitator for e-partnerships,' *Proceedings of the 5th International Conference on Autonomous Agents*, Montreal, Canada, pp. 108-109.

[51] Orlikowski, W. (1993). 'Learning from Notes: Organizational Issues in Groupware Implementation', *The Information Society*, Vol. 9, No. 3, pp. 237-250.

[52] Pancerella, C. M., Rahn, L. A., and Yang, C. L. (1999). 'The diesel combustion collaboratory: combustion researchers collaborating over the Internet,' *Proceedings of the 1999 ACM/IEEE Conference on Supercomputing (CDROM)*, Portland, Oregon, USA.

[53] Rapoport, A., and Guyer, M. (1966). A taxonomy of 2 X 2 games, *General Systems*, Vol. 2, pp. 203-214.

[54] Ratnasingham, P., and Kumar, K. (2000). 'Trading partner trust in electronic commerce participation,' *Proceedings of the 21st International Conference on Information Systems*, Brisbane, Queensland, Australia, pp. 544-552.

[55] Sairamesh, J., Mohan, R., Hason, C., and Bender, J. (2002). 'A platform for business-to-business sell-side, private exchanges and marketplaces,' *IBM Systems Journal*, Vol. 41, No. 2, pp. 242-254.

[56] Scott, S. C., and Lane, V. R. (2000). 'A stakeholder approach to organizational identity,' *Academy of Management Review*, Vol. 25, pp. 43-62.

[57] Semich, J. (1994). 'Information replaces inventory at the virtual corporation,' *Datamation*, Vol. 40, No. 14, pp. 37-40.

[58] Sheppard, B. H., and Tuchinsky, M. (1996). 'Micro-OB and the network organization,' In R. M. Kramer and T. R. Tyles (Eds.), *Trust in organizations: frontiers of theory and research*, Thousand Oaks, CA: Sage, pp. 140-165.

[59] Snow, C., Miles, R., and Coleman, Jr. H. (1992). 'Managing the 21st century network organizations,' *Organizational Dynamics*, Vol. 20, Winter, pp. 5-19.

[60] Staples, S. D., and Ratnasingham, P. (1998). 'Trust: the panacea of virtual management?' *Proceedings of the International Conference on Information Systems*, Helsinki, Finland, pp. 128-144.

[61] Steinfield, C., Jang, C. Y., and Pfaff, B. (1999). 'Supporting virtual team collaboration: the TeamSCOPE System,' *Proceedings of the International ACM SIGGROUP Conference on Supporting group work*, Phoenix, Arizona, pp. 81-90.

[62] Swagerman, D. M., Dogger, N., and Maatman, S. (2000). 'Electronic markets from a semiotic perspective,' *Electronic Journal of Organizational Virtualness*, Vol. 2, No. 2, pp. 22-42. <http://www.virtual-organization.net>.

[63] Thibaut, J., and Kelley, H. H. (1959). *The Social Psychology of Groups*, Wiley and Sons, New York.

[64] Vandenbosch, B., and Ginzberg, M. (1997). 'Lotus Notes and Collaboration: Le Plus Ca Change...', *Journal of Management Information Systems*, Vol. 13, No. 3, pp. 65-81.

[65] Venkatraman, N., and Henderson, J. C. (1998). 'Real strategies for virtual organizing,' *Sloan Management Review*, Fall, pp. 33-48.

[66] Wachter, H., and Reuter, H. (1992). 'The ConTract Model,' In *Database Transaction Models for Advanced Applications*, edited by A. Almagramid, Morgan Kaufmann.

[67] Whittaker, S. (1996). 'Talking to strangers: an evaluation of the factors affecting electronic collaboration,' *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, Boston, Massachusetts, pp. 409-418.

[68] Wicks, A. C., Berman, S. L., and Jones, T. M. (1999). 'The structure of optimal trust: moral and strategic implications,' *Academy of Management Review*, Vol. 29, pp. 99-116.

[69] Wiesenfeld, B. M., Raghuram, S., and Garud, R. (1998). 'Communication patterns as determinants of organizational identification in a virtual organization,' *Journal of Computer Mediated Communication*, Vol. 3, No. 4, pp. 1-22.

[70] Wigand, R. T. (1977). 'A Cybernetic Model of Communication and Interorganizational Relationships among Complex Organizations,' *Proceedings of the Eighth International Congress on Cybernetics*, 8, 463-480.

[71] Wigand, R. T. (1978). 'A Dynamic Model of Interactions among Complex Organizations within an Organization-set,' In R. Trappl et al., (eds.), *Progress in Cybernetics and Systems Research*. Washington, DC: Hemisphere Publishing, Vol. 5, 32-37.

[72] Wigand, R. T., Picot, A., and Reichwald, R. (1997). *Information, organization, and management: Expanding markets and corporate boundaries*, Chichester, UK: Wiley

Trust Models in the E-Commerce Environment

Siddhi Pittayachawan, Mohini Singh

School of Business Information Technology, RMIT University
GPO Box 2476V, Melbourne, Victoria 3001, Australia
siddhi.p@ieee.org, mohini.singh@rmit.edu.au

ABSTRACT

This paper analyses several security systems and aggregates their characteristics supporting trust. These characteristics are then matched with e-business models to try and identify the most suitable security system for each model. This is preliminary work undertaken to establish appropriate trust models in the e-commerce environment. The models discussed in this paper are hypothetical.

Keywords: electronic commerce, public key infrastructure, trust, trust model

1. INTRODUCTION

The Internet is the largest global network supporting e-business models, B2B (business-to-business), B2C (business-to-consumer), C2C (consumer-to-consumer), or G2C (government-to-consumer) e-commerce. E-businesses use the Internet to provide product information, online catalogue, electronic transactions, business exchanges, e-negotiations, e-procurement, and online services (e.g. e-government, e-banking and e-insurance) to customers and business partners. New and evolving technologies can be combined with the Internet to enhanced business services. Business organizations around the world are capitalising on the Internet to expand business to organizations and customers to greater geographic regions. Turban et al. [27] emphasise that the benefits of e-businesses include reduced costs, automated and integrated business processes, quick retrieval and dissemination of information, better information management methods, and efficient transactions. In the B2B e-commerce, e-procurement is an important application enabling organisations to reduce purchasing administrative costs by a substantial amount. In the B2C, e-business consumers have the advantage of a 24 by 7 shopfront, access to e-banking and e-government services. [27]. The SMEs (Small and Medium Enterprises) are also capitalising on the Internet to compete with larger businesses.

Although security technologies available off the shelf, these are not always sufficient to prevent the Internet from various kinds of attacks (virus, information access, worms, etc.), an important barrier to e-commerce has been a lack of trust by the Internet users to complete a transaction. Trust has been found to be a crucial factor for e-commerce success [20] due to uncertainty and risk in its nature. As e-commerce environment becomes more uncertain because Internet users are separate from each other by space (e.g. distance between countries) and time (e.g. delayed response on the Internet), the need for trust is very critical [10, 20].

This paper introduces and discusses trust models suitable for different e-business models (e.g. B2C, B2B, B2E, and G2G). These models were developed by analysing the trust mechanisms from several existing trust models, including ITU-T Recommendation X.509 [1], PEM (Privacy Enhanced Mail) [13], PGP (Pretty Good Privacy) [2], PEMToolKit [3], ICE-TEL (Interworking public key Certification infrastructure for Europe) [6], SDSI (Simple Distributed Security Infrastructure) [23], SPKI (Simple Public Key Infrastructure) [8], NPKI (Nested certificate based PKI) [17], and Solar Trust Model [7]. Note that SDSI and SPKI has been merged and become one trust model as the version 2 of SDSI/SPKI recommendations were published in 1997, due to the similarity of these two models. These eight trust models are PKI (Public Key Infrastructure) -based models from either international security standards for IT (Information Technology), purposed models in published paper, or existing models that have been successfully deployed in several business applications.

2. TRUST

Technology trust has been studied for at least half a century. These studies have included meanings [19], characteristics [20], calculation of trustworthiness [18], and the relationships between trust and other factors, such as risk, uncertainty, and confidence [14, 16, 24, 26, 28]. The impact of trust on the human society, business and commercial partners, organisations, and teamwork has also been investigated. Regrettably, to date there is no satisfactory explanation of the nature of trust or its relationship to these entities [24]. As a result, the meanings and characteristics of trust in many technical applications are still imprecise. McKnight and Chervany [20] identified sixteen distinctive categories of trust characteristic definitions (competent, expert, dynamic, predicable, good and moral, good will, benevolent and caring, responsive, honest, credible, reliable, dependable, open, careful and safe, shared understanding, and personally attractive) grouped into five major categories of competence, predicability, benevolence, integrity, and other. Their findings demonstrated that trust is a

relationship that can be seen and used from several perspectives. In this paper, trust refers to one's belief of others in terms of competence, predicability, benevolence, or integrity in the e-commerce environment.

3. TRUST IN E-COMMERCE

The reason why trust has become a very important issue in e-commerce is that the environment and digital processes (e.g. electronic transactions) of e-commerce contain very high risk factors, such as impersonation, fraud, security, privacy, dishonest people, page-jacking, and parallel webs [5, 14, 16, 20, 25, 26]. Hoffman, Novak, and Peralta are of the opinion that almost 95 percent of online users decline to provide personal information on web sites due to a lack of trust [12]. They also suggest that 69 percent of online users did not provide information on web sites because the sites did not provide any information on how the data would be used. Ponemon Institute (<http://www.ponemon.org>) and TRUSTe (<http://www.truste.com>) reported that 76 percent of Internet users are concerned about "identity theft" if their personal information were leaked to unauthorised individuals or organisations [11]. Grazioli and Jarvenpaa emphasise that there are approximately 25 million pages, or 2 percent of the total number of pages on the Internet supporting fraud, called "page-jacking"[10].

Trust is important wherever risk, uncertainty, or interdependence exists [20]. Without trust, e-commerce will not be a success [26]. It is one of the most desired qualities in any close relationship. It is indispensable in social relationships, which may lead to significant benefits especially in business relationships [14]. Trust reduces complexity in human society [16]. Similarly, it is a bridge for both a seller and a buyer to cross over uncertainty in the e-commerce environment. Trust problems affect family relationships, business transactions, and client/professional interactions [28]. A buyer wants to buy a quality product with a reasonable price while a seller wants to sell a product and to be well known in the marketplace. In fact, a buyer could be a fraudster or a seller could sell a non-qualified product - or nothing at all in the e-commerce environment.

Before e-commerce had been established, there was only one type of commerce called "brick-and-mortar commerce." In the marketplace, products could be seen, touched, and tested at the point of sale. Tan and Thoen [26] suggest that it is difficult to increase the trust of online users in e-commerce as compared to brick-and-mortar commerce because buyers and sellers cannot see each other and someone could impersonate somebody else, either known or non-existent. This makes trust in the online environment very vulnerable. E-commerce is known for receiving payment and not sending the product to the buyer. This occurred with eBay many

times. Auction fraud had increased from 106 cases in 1997 to 25,000 cases in 2001 [22].

Another type of risk perceived by the online shopper is losing control over the situation and/or not being familiar with this kind of technology. "Social uncertainty" exists when the seller has an incentive to act in a way that imposes cost or harm on the buyer, and the buyer does not have enough information to predict the behaviour of the seller [10].

4. EVOLUTION OF TRUST MODELS

In 1976, Diffie and Hellman [8] introduced the PKCS (Public Key Cryptosystem), which is a cryptography method, a central authority or public file to support e-mail security. This scheme reduced the risk of key management, which is a method of managing a key pair that consists of a public key and a private key. However, with this method, an impersonation was still possible because no one could ensure that the public key that online users obtained from the trusted public directory really belonged to the claimed entity.

In 1978, Kohnfelder invented the idea of a digital certificate [15]. It was a mechanism designed to link the public key, which is a tool to encrypt a plain message and can be opened by the owner of that key, to a given identity, and signed by a trusted entity such as TTP (Trusted Third Party). Depending on the method of encryption used, the digital certificate could be almost unforgeable, or take a long time to be deciphered. This method solved the impersonation issue previously mentioned and improved the performance of key management for the TTP [8].

In 1988, the CCITT (Comité Consultatif Internationale de Telegraphie et Telephonie), which is now known as the ITU (International Telecommunication Union), published CCITT Recommendation X.509. Part of X.509 was to define and standardise a global, distributed database of named entities, such as people, computers, printers, etc. It also could be described as an online telephone book. However, the plan was not a success because the idea of using a single global name in the world that had countless number of entities was unlikely to be true [8].

In 1989, PEM (Privacy Enhanced Mail) attempted to implement the X.509 standard by the IETF (Internet Engineering Task Force). However, it was delayed due to the long time spent on deploying its infrastructure, including IPRA (Internet Policy Registration Authority), PCA (Policy Certificate Authority), and CA (Certificate Authority) [6].

In 1991, Zimmermann [2] introduced new secure-communication software known as PGP (Pretty Good Privacy). The structure of PGP was different from X.509 and PEM. Unlike PEM that had to wait for the

establishment of a single global root and a hierarchy of CAs, PGP allowed a digital certificate to be signed by anyone, and could contain multiple digital signatures. This approach enabled several virtual communities to be quickly established and grown due to the “Six Degrees of Separation” theory, which describes how someone can connect to anyone in the world through the chain of intermediaries containing not more than six people [4], and was well known as the “web of trust” model.

In 1992, the NSF (National Science Foundation) enabled commercial companies to conduct business transactions securely over the Internet. With the establishment of this large global network, many companies lodged business online. However, the Internet was not suited for a commercial environment and was not developed with security in mind [14]. It was meant for sharing information in plain text format.

5. TRUST MODEL ISSUES

A trust model should be able to support trust relationships that are required by users and online businesses, and to provide control mechanisms that allow them to establish and enhance trust. Therefore, it is important to understand the characteristics and needs of target community and users, and to create and embed these characteristics into a trust model [6]. The framework of a trust model is an important factor to determine how the model will be used and whether it is suitable for the target virtual community. The framework of a trust model in this paper refers to trust mechanisms to manage trust relationships between buyers, sellers, suppliers and other relevant parties. If the target community is a small group of casual end-users but a trust model uses a very strict security policy, then model deployment, user registration, and cross certification will be very difficult and slow to manage. This happened with PEM that contains very strict security policy and requires deploying several central authorities before any user is able to communicate with each other securely. On the other hand, if the target community is made up of a large number of end-users and CAs but the trust model lacks a standard security policy, then that virtual community will not be able to function successfully. PGP can be associated with this characteristic because it contains no standard security policy, and therefore, it is not easily scalable when hundreds of thousands of users are involved [6].

Although a trust model is not only a security system [14, 16], in this paper, it is based on the analysis of PKI-based security systems. Security system in e-commerce is different from security in traditional networks. There are four major security issues in e-commerce [25]:

- Authentication – communicating parties must be certain of each other’s identity and/or credentials;
- Confidentiality – data must not be visible to eavesdroppers;

- Integrity – communicating parties must know when data has been tampered with; and
- Non-repudiation – it must be possible to prove that a transaction has taken place.

However, Skevington [25] argued that this approach is inadequate in the open and distributed environment of the Internet. Trust must be embedded into infrastructure, data, and user identity.

6. RESEARCH METHODOLOGY

The research discussed in this paper was inspired by a lack research on trust mechanisms illustrating how trust models perform or apply to e-commerce business models. It is based on a document analysis methodology. The findings in this paper were compiled using a categorical aggregation analytic strategy. Categorical aggregation is the process of piecing together bits of information gathered regarding an issue and organising it into an orderly research interpretation [29]. Each of these categories is further broken down into nominal attributes (non-numeric and unordered elements) using both homogeneous and heterogeneous decomposition methods. Goldstein and Roth [9] stated that there are two types of decomposition. Homogeneous decomposition is to use the same attribute to repeatedly partition a group by choosing more narrow ranges of the attribute’s values. Heterogeneous decomposition, on the other hand, is to use different attribute to decompose sets for successive partitions.

7. FINDINGS

An analysis of eight trust models chosen for this research revealed that some of the findings are common to all while some are quite different. Trust models examined in the research are all PKI-based. The research found that there are 25 trust mechanisms in the chosen trust models. Although the value or importance of trust mechanisms were not identified, these trust mechanisms were ranked based on the opinions of [2, 3, 6-8, 17, 21, 23] who previously stated in their works, the commonality of trust models, and the effect of these mechanisms on communities when attributes were changed. In this paper, an attribute is a sub-category of a trust mechanism. Table 1, which is presented at the conference, shows trust mechanisms and their attributes in the trust models commonly used. The horizontal top column shows the name of each trust model. The vertical left-most column lists trust mechanisms. Each trust model uses different attributes to create. Some trust models support more attributes than others. This illustrates that they are more flexible than other models. However, it does not mean that they can support more kinds of e-commerce models. An appropriate combination of trust mechanisms is crucial in order to develop a trust model suitable for each kind of e-commerce models.

8. TRUST MODELS

In order to determine the most appropriate trust model for an e-business model, trust mechanisms are matched with the interaction between participants. The behaviour of a trust model may affect how users interact with each other, what kind of information (e.g. user's identity, security policy) is available, how information can be validated or delivered, what kind of environment is used (e.g. user-friendly interface, control mechanisms), and for which e-commerce model is suitable. Trust mechanisms are used to assist in the development of the framework of a trust model for each e-commerce model.

In this section, seven conceptual trust models are discussed. These models were developed from 25 trust mechanisms presented in Table 1. The combinations of attributes of trust mechanisms are based on the application and the type of network in which the hypothetical e-commerce model operates. The models are discussed from the smallest community containing the simplest trust relationships (e.g. relationships between friends, family, relatives, and team-mates), to the largest community containing very complex trust and several kinds of trust relationships (e.g. relationships between business partners, companies, government, organisations). The model starts from the simplest trust relationships or close relationships.

8.1 Close-Relationship Trust Model (B2E, B2C)

The simplest human society starts from close relationships that are established directly between relevant parties (PGP Corporation 2003). It contains the highest level of trust because members in this community know each other very well. It also needs the least secure environment, which makes this trust model very simple, compared to other sample trust models in this section. Figure 1, which is presented at the conference, illustrates a small group of end-users that have agreed to establish an internal network with a convenient communication method based on an adequate security system.

This type of trust model could be used for private trust relationships such as friends, members in the family, relatives, colleagues, or a group of people that know each other well, and this makes the community small. Thus, there is no need for any formal proof signed by some trustworthy entities. It may not need a strict security policy unless there is some very sensitive information that needs to be shared among a few exclusive members in the same community. In the e-commerce context, it may apply to a small B2C enterprise that contains single security domain or B2E (Business-to-Employee) e-commerce, an Ethernet, or the requirement of establishing a secure communication channel (e.g. e-mail).

Members in this small community are assumed to be familiar with each other and do not need a central authority because it is a close domain. Therefore, the structure is anarchy, growth is organic, trust management is decentralised, and trust relationships are managed by a trusted entity. There is no standard security policy and trust transitivity in this model because members do not need to contact people in other communities. From Table 1, it can be said that PGP is most suitable for this e-business model.

8.2 Casual Trust Model (B2C, B2B)

A casual trust model in this paper refers to the illustration in Figure 2 (presented at the conference) when:

- Users need to contact other users, who are members in different security domains;
- Users need to create a central authority for enabling a standard security policy for user authorisation system; and/or
- Users need to strengthen the security level throughout their community.

With any of these reasons, a security policy must be standardised and this means that a community needs to establish a central authority. In this paper, a central authority is a person who is responsible for validating the identity of members, signing digital certificates, creating security policy, issuing cross certificates, and maintaining the network. Thus, this trust model is suitable for a medium enterprise that contains a few different security domains. In B2B e-commerce, two companies need to establish security communication channels between several security domains (e.g. sending or sharing sensitive information between two companies). In B2C e-commerce, a customer needs to do business with a company.

Members in this community need to contact people in other communities, presumably in an e-commerce environment containing a few companies and customers. Security policy is needed to properly standardised in order to enable cross certification, which is a task done by a central authority. Therefore, the structure is both hierarchy and anarchy, growth is both scalable and organic, trust management is both centralised and decentralised, trust relationships are established by either a trusted entity or a trusted path. These cover security policy and trust transitivity. From Table 1, it can be said that ICE-TEL is the most suitable for this e-business model.

8.3 Community Trust Model (B2C, B2B, G2C, G2G)

A community trust model is used when:

- Users need to establish proper formal small communities on the Internet;
- Organisations need to set up a trustworthy network for secure communication; or

- Companies want to create their communities in order to deploy an e-commerce environment and applications.

In this model, there would be more than one type of trust relationship because it is an open community, which anyone could join in. This does not limit them to only a close-relationship as it does in the trust models previously discussed. Possible types of trust relationships may include close relationship, acquaintance, friends, family, colleagues, co-workers, customers, business partners and other distant relationships (e.g. third parties). An appropriate set of security policies is needed in order to prevent hackers and malicious attacks. This trust model is suitable for an open community where there are several kinds of companies and organisations in the same environment, e.g. B2B or G2G exchanges of goods between large and small organizations or suppliers and distributors, or B2C or G2C e-commerce where customers need to filter unrelated companies and measure the trustworthiness of related companies.

A number of e-commerce organisations are large and contain different sizes of companies and businesses. Members in this community need a more sophisticated trust model for trust management. A suitable structure for this will be both hierarchy and anarchy, scalable and organic growth with trust management both centralised and decentralised. From Table 1, it can be said that ICE-TEL is suitable for this e-business model.

8.4 Community with Casual Trust Model (B2C, B2B, G2C, G2G)

This model has the same characteristics as the community trust model with an extra environment for users to create their own private communities. It is actually a casual community inside the community trust model. This means that this trust model contains a community trust model as a primary model, and a casual trust model, as a secondary trust model. Therefore, users can create casual communities, and save time and bandwidth requirements in the process of verifying identities. This model is suitable for a larger scale community compared to the previous models. For example, a virtual community where there are several companies, governments, and users in the same environment; it will incorporate central system administration domain, and share sensitive information between several member companies in the same department that contains members from different companies or organisations. Another example is that a large company that contains several large divisions or sub-companies, and wants to create a new special division that contains a few selected members from different divisions.

For this, bottom-up virtual society establishment, which is a unique feature of PEMToolKit, is chosen. From

Table 1, it can be said that ICE-TEL is the most suitable for this e-business models.

8.5 Organisational Trust Model (G2G)

An organisational model is to be used for a large organisation having a solid structure that is unlikely to be changed, or a community that contains very sensitive information and needs very secure communication channels. One of the most suitable business models is G2G (Government-to-Government). Hence, it is crucial to not only to increase trust but also reduce risk in a community. Members in this community need a very secure and solid model. Trust relationships established by users are not allowed in order to minimise risk. In fact, this community seeks more security rather than trust. From Table 1, it can be said that X.509 is the most suitable for this e-business model.

8.6 Popularity-based Trust Model (C2C, P2P)

This model is based on the popularity of each user. It may be used with a measurement of trustworthiness of users in some closed communities, such as an online auction, bookshop, or e-commerce company search engine. For example, it would be better if an online auction could provide trustworthy information about buyers or sellers before transactions have been processed. The trustworthiness value would be determined by how many digital signatures have been signed on the digital certificate of the target and who signs those digital signatures.

In this model, if a member registers online, then a central authority would not sign on a digital certificate because it is too difficult to trust and verify all information provided digitally from a faceless member. However, if a member registers at a physical office, then a central authority would verify proof of identity (e.g. driving license, social security number, passport or other personal id cards). A central authority would then help one to generate a key pair with a digital certificate signed by a central authority. This process is long but it counterbalances the problem of new members having no trustworthiness information. The local business registrar should be responsible for this task, as all companies must be registered with the government organisations before commencing any business. However, a central authority could revoke the digital signature for any member if there is any suspicion. Therefore, trust relationships in this trust model would be changed dynamically by comparison with other trust models. This trust model is suitable for a large community that contains members with the same level of authorisation and needs the value of trustworthiness or background information of users in order to decide which one is trustworthy enough to do the business. One of the suitable business models using popularity-based trust model is C2C (Consumer-to-Consumer), P2P (Peer-to-Peer) or auction-based e-commerce environment.

Figure 7, which is presented at the conference, illustrates how this model is superior to other models previously discussed in this section. Members in this closed community need a dynamic and flexible trust relationship. Therefore, structure is anarchy, growth is organic, and trust management is decentralised. From Table 1, it can be said that PGP is the most suitable for this e-business model.

8.7 Integrated Trust Model

This trust model is the combination of casual, community, organisational, and popularity-based trust models. It is a trust model, which could support different kinds of trust relationships in the same or across communities. This trust model is suitable for a very large and complex community that contains many relationships. In addition, both central authorities and end-users are included. End-users are also able to create their own communities in order to create either open or closed communities. In fact, this model will be used when all models above have been already deployed and users need to standardise or unify their communities in order to establish secure communication channels conveniently. Suitable e-business models that can use this model are B2C, B2B, G2C, and G2G when calculation of trustworthiness is important.

An integrated trust model firstly uses a hierarchical structure as a backbone in order to properly define and distribute a set of security policies for different security domains. If a community is very large and contains different types of members, then an IPRA (Internet Policy Registration Authority) and PCAs (Policy Certificate Authorities) may be needed. The part involving an IPRA to a CA is an organisational trust model, and the part involving a CA to an end-user is a community trust model. From Table 1, it can be said that ICE-TEL is most suitable for this e-business model.

9. CONCLUSION

In this paper, a first attempt has been made to match appropriate trust models to e-commerce models for managing trust. Further work will be done to prove the discussion in this paper.

REFERENCES

[1] ITU-T Recommendation X.509, "Information technology - Open systems interconnection - The Directory: Public-key and attribute certificate frameworks," March, 2000
 [2] "An Introduction to Cryptography," PGP Corporation, June 2004 <http://www.pgp.com> [Accessed 28 September 2004]
 [3] Bahreman, A., "PEMToolKit: Building a Top-Down Certification Hierarchy for PEM from the Bottom Up," presented at Proceedings of the 1995 Symposium

on Network and Distributed System Security (SNDSS), San Diego, California, USA, 1995.

[4] Blass, T., "Stanley Milgram," 29 September 2003 <http://www.stanleymilgram.com> [Accessed 9 January 2004]
 [5] Castelfranchi, C. and Y.-H. Tan, "The Role of Trust and Deception in Virtual Societies," presented at Proceedings of the 34th Hawaii International Conference on System Sciences (HICSS), Maui, Hawaii, USA, 2001.
 [6] Chadwick, D. W., A. J. Young, and N. K. Cicovic, "Merging and Extending the PGP and PEM Trust Models - The ICE-TEL Trust Model," in *IEEE Network*, vol. 11, 1997, pp. 16-24.
 [7] Clifford, M., C. Lavine, and M. Bishop, "The Solar Trust Model: Authentication without Limitation," presented at Proceedings of the 14th Annual Computer Security Applications Conference (ACSAC), Scottsdale, Arizona, USA, 1998.
 [8] Ellison, C. M., B. Frantz, B. Lampson, R. Rivest, B. Thomas, and T. Ylonen, IETF RFC 2693, "SPKI Certificate Theory," September, 1999
 [9] Goldstein, J. and S. F. Roth, "Using Aggregation and Dynamic Queries for Exploring Large Data Sets," presented at Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence, Boston, Massachusetts, USA, 1994.
 [10] Grazioli, S. and S. L. Jarvenpaa, "Perils of Internet Fraud: An Empirical Investigation of Deception and Trust with Experienced Internet Consumers," in *IEEE Transactions on System, Man, and Cybernetics - Part A: Systems and Humans*, vol. 30: IEEE, 2000, pp. 395-410.
 [11] Greenspan, R., "Consumers Trust eBay," Ecommerce-Guide.com, 16 June 2004 <http://www.ecommerce-guide.com> [Accessed 14 July 2004]
 [12] Hoffman, D. L., T. P. Novak, and M. Peralta, "Building Consumer Trust Online," in *Communications of the ACM*, vol. 43, 1999, pp. 80-85.
 [13] Kent, S., IETF RFC 1422, "Privacy Enhancement for Internet Electronic Mail: Part II: Certificate-Based Key Management," February, 1993
 [14] Kini, A. and J. Choobineh, "Trust in Electronic Commerce: Definition and Theoretical Considerations," presented at Proceedings of the 31st Hawaii International Conference on System Sciences (HICSS), Kohala Coast, Hawaii, USA, 1998.
 [15] Kohnfelder, L. M., "Towards a Practical Public-Key Cryptosystem," in *Department of Science: Massachusetts Institute of Technology*, 1978, pp. 51.
 [16] Konrad, K., G. Fuchs, and J. Barthel, "Trust and Electronic Commerce - More Than a Technical Problem," presented at Proceedings of the 18th IEEE Symposium on Reliable Distributed Systems (SRDS), Lausanne, Switzerland, 1999.
 [17] Levi, A. and M. U. Caglayan, "An Efficient, Dynamic and Trust Preserving Public Key Infrastructure," presented at Proceedings of the 2000

IEEE Symposium on Security and Privacy (S&P), Berkeley, California, USA, 2000.

[18] Marsh, S. P., "Formalising Trust as a Computational Concept," in *Department of Computing Science and Mathematics*: University of Stirling, 1994, pp. 163.

[19] McKnight, D. H. and N. L. Chervany, "The Meanings of Trust," MIS Research Center, University of Minnesota, 1996 [Accessed 29 September 2004]

[20] McKnight, D. H. and N. L. Chervany, "Conceptualizing Trust: A Typology and E-Commerce Customer Relationships Model," presented at Proceedings of the 34th Hawaii International Conference on System Sciences (HICSS), Maui, Hawaii, USA, 2001.

[21] Perlman, R., "An Overview of PKI Trust Models," in *IEEE Network*, vol. 13, 1999, pp. 38-43.

[22] Reuters, "Read this before buying on eBay," Lycos, Inc., 9 January 2003 <http://www.wired.com> [Accessed 28 January 2003]

[23] Rivest, R. L. and B. Lampson, "SDSI - A Simple Distributed Security Infrastructure," 30 April 1996 <http://www.syntelos.com/spki> [Accessed 28 September 2004]

[24] Robles, S., S. Poslad, J. Borrell, and J. Bigham, "Adding Security and Privacy to Agents Acting in a

Marketplace: A Trust Model," presented at Proceedings of the 35th International Carnahan Conference on Security Technology (ICCST), London, England, 2001.

[25] Skevington, P. J., "From Security to Trust - Creating Confidence to Trade Electronically," presented at Proceedings of the IEE Colloquium on eCommerce - Trading But Not As We Know It, London, England, 1998.

[26] Tan, Y.-H. and W. Thoen, "Formal Aspects of a Generic Model of Trust for Electronic Commerce," presented at Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS), Maui, Hawaii, USA, 2000.

[27] Turban, E., D. King, J. K. Lee, and D. Viehland, *Electronic Commerce 2004: A Managerial Perspective*, 3rd Edition ed: Prentice Hall, 2004.

[28] Webb, W. M. and P. Worchel, "Trust and Distrust," in *The Social Psychology of Intergroup Relations*, W. G. Austin and S. Worchel, Eds. Monterey, California, USA: Brooks/Cole Publisher, 1979, pp. 213-228.

[29] Wilson, B. B. and D. L. Stewart, "Employers' Perceptions of Welfare Reform: Implications for Cooperative Extension Personnel," *Journal of Extension*, vol. 38, 2000.

A Managers' Guide to KMS Adoption and Diffusion

Jun Xu¹, Mohammed Quaddus²

¹Graduate College of Management, Southern Cross University, Australia

²Graduate School of Business, Curtin University of Technology, Australia
jxu@scu.edu.au; quaddusm@gsb.curtin.edu.au

ABSTRACT

This paper reports and summarises the important findings of a study into the factors that influence the adoption and diffusion of knowledge management systems (KMSs) in Australian organizations. The implications for managerial practices are discussed. The future research directions are also presented.

Keywords: Research Summary, Managerial Implications

1. INTRODUCTION

There has been a major shift towards a knowledge economy over the past decade. Coupled with extreme competition, shorter product life-cycles, and rapid development in technology, organizations are paying more attention to safeguarding their knowledge-based assets. Knowledge is now considered the most important organizational asset for obtaining sustainable competitive advantage (Wiig, 1997). Organizations are paying more attention to organizational knowledge and KM. They are looking at ways to maximize their knowledge-based assets for the purpose of staying ahead of the competition. Although knowledge and knowledge management (KM) are not new concepts, knowledge management systems (KMSs), which involve the application of IT systems and other organizational resources to manage knowledge strategically in a more effective and systematic way (Alavi & Leidner, 1999), are a relatively recent phenomenon. While a KMS (or some version of it) is widely applied in organizations, the topic has not been well explored by researchers and scholars in an empirical way. There is a scarcity of empirical studies of KMSs, especially in the area of their adoption and diffusion. This paper reports the findings of an Australian study of KMS adoption and diffusion in organizations. This study was done in a comprehensive manner and findings have been reported elsewhere on different aspects of the study (Quaddus, Xu & Wood, 2002; Quaddus and Xu, 2002; Quaddus and Xu, 2003; Xu, Quaddus & Wood, 2001). This paper presents a summary of the findings and articulates the managerial implications of them.

The paper is organized as follows. The capstone summary of the study findings is presented in section 2. Section 3 presents the comprehensive managerial implications of the results. The conclusions and future directions are presented in section 4.

2. CAPSTONE SUMMARY OF THE STUDY FINDINGS

This research attempts to get answers for the research question: What are the factors that influence the adoption and diffusion of knowledge management systems (KMSs) in Australian organizations? In doing so it identifies the factors affecting the diffusion of KMSs based on selected Australian companies at both the organizational level (identifying what causes organizations to adopt and implement a KMS) and the individual level (identifying what causes individuals/end users to accept or reject a KMS). The study also focuses on identifying the relationships between the external influences, the perceptions of the KMS, and the diffusion of KMSs in Australian organizations to answer the second research question: What must be done for the successful diffusion of a KMS? The research proposes and examines a model of KMS adoption and diffusion developed from both the literature review and practical field studies. The details of the study can be found in Xu, Quaddus & Wood, 2001; Quaddus, Xu & Wood, 2002; Quaddus & Xu, 2002; Quaddus & Xu, 2003; and Xu, 2003. The research has resulted in a number of findings to explain the adoption and diffusion of KMSs in organizations:

1. Perceived usefulness has a positive relationship with an organization embarking on a KMS. Perceived user-friendliness has a significant effect on the initiation of a KMS indirectly through perceived usefulness. Perceived usefulness is the primary driver of an organization's interest in seeking a KMS to manage its knowledge in a better and more effective way.
2. People's decision to accept and use a KMS is directly determined by perceived user-friendliness and perceived voluntariness jointly. Meanwhile, perceived usefulness also has a significant effect on people's decision to use the system. But user-friendliness has the greatest total effects on an individual's intention to use the system, followed by perceived usefulness, and then perceived voluntariness.

3. The impact of individual differences and task complexity on the organization's decision about KMS adoption and diffusion is mediated by the perceived usefulness.
4. The influence of individual differences and task complexity on people's use of a KMS is mediated by the perceived usefulness.
5. The KMS diffusion process consists of six stages: initiation, adoption, pilot implementation, organic growth, organizational implementation, and diffusion/sustained use of KMS.

The details of the results can be found in Xu, Quaddus & Wood, 2001; Quaddus, Xu & Wood, 2002; Quaddus & Xu, 2002; Quaddus & Xu, 2003; and Xu, 2003.

3. MANAGERIAL IMPLICATIONS

The results of this study have significant implications for managerial practices. The managerial implications arising from this study are discussed from various perspectives:

- ◇ The need for a KMS
- ◇ Approach to implementing a KMS
- ◇ Cost of a KMS
- ◇ The comprehensive model of KMS adoption and diffusion
- ◇ Significant factors of KMS adoption and diffusion
- ◇ End-user focus and involving people in the KMS
- ◇ Organizational adjustment to embrace the KMS
- ◇ Role of top management in KMS implementation
- ◇ Technology for the KMS
- ◇ Suggestions for KMS adoption and diffusion

The Need for a KMS

It is necessary for organizations to adopt and implement a KMS to survive competition and gain competitive advantage in the knowledge economy. Organizations need better control of the knowledge they already have. Many organizations do not know what they have. KMSs have the ability to make an organization better and help individuals do their jobs better. A KMS can make a contribution in the following areas (see Quaddus & Xu, 2002):

- § Help decision-makers cope with challenges and problems.
- § Make people more effective in their jobs.
- § Inspire people to become more creative in their jobs.
- § Make organizations more productive.
- § Cut operating costs and time for production and product development by avoiding the need to reinvent the wheel.
- § Increase people's knowledge and enlarge their knowledge base.
- § Avoid repeating the same mistakes.

- § Provide better service to customers.

Approach to a KMS

A KMS is a broad way or approach to deal with the management of both tacit and explicit knowledge. This notion is supported by the majority of the survey respondents (see Quaddus & Xu, 2002). Many organizations do not have a KMS even though they have a number of systems that are consistent with a KMS. But they have not put the whole package together. A KMS consists of a number of elements, such as structure, culture, job descriptions, information technology (IT), processes, procedures, etc. A KMS is also a multi-process. It actually touches every part of the business and everything the organization does. At the same time, IT can be an important enabler to support KM in various ways. The KMS is a much broader subject than IT. It is a way, mechanism and structure to manage organizational knowledge. Organizations should focus on building a knowledge-sharing culture and structure as well as people's capabilities before they adopt or build a system.

The first necessary and important step for setting up an effective KMS is to review the way an organization operates and determine the easiest method of collecting and transferring the important types of knowledge, which will provide competitive advantage (Mazzei, 2000). The organization should understand what knowledge means to its business and find out what the potential impact of the knowledge will be. It has to identify what might be worth doing and where the leverage points might be. For example, the most important knowledge domains identified in this study were knowledge of customer services, knowledge of strategic planning, knowledge of marketing sales, knowledge of competition (see Quaddus & Xu, 2002), which implies that when an organization is developing its KMS, it should include this knowledge into its KMS and provide cost-effective access to these knowledge areas. Furthermore, the cost-benefit analysis of a KMS can also be linked to organizational performance in the corresponding perspectives, such as customer satisfaction, business strategy development, product development and innovation, time to market, product promotion, market share, competitive positioning, and so on (Alavi & Leidner, 1999). The results of this study suggest that a KMS has the capability to provide both external knowledge (such as customer information, competitive information, market information, business partner information, and supplier information) and internal knowledge (such as operations information, human resources information, financial and accounting information) to people. That will help them figure out how to structure, allocate responsibilities and draw the process through the business. The first thing in KM is addressing the different areas of knowledge content and business processes relating to knowledge that are critical to the business. Then the organization should

adjust its business processes and organizational structure accordingly to embrace the KMS.

Cost of a KMS

In the survey, the respondents reported the total cost of KMSs in their organizations up to the time the questionnaire was answered as ranging from less than AUST\$100,000 to more than AUST\$1,000,000 (see Quaddus & Xu 2002). However, these figures should be interpreted with caution since there are difficulties in accounting for building, operating and maintaining a KMS, and also in measuring the system benefits (Sarvary, 1999). Before organizations put KMSs into place, they have to look at their corporate infrastructure (Brooking, 1999). Corporate infrastructure includes a broad range of infrastructure assets, such as management philosophy, corporate culture, management and business processes, financial relationships, methodologies, and IT systems (Brooking, 1999). The investment on a KMS should cover both technological infrastructure (i.e., KMS technologies – both hardware and software, MIS department, etc) and organizational infrastructure (i.e., user training, rewards for knowledge sharing, changing organizational culture, structure, etc).

A Comprehensive Model of KMS Adoption and Diffusion

This research presents a practical model of KMS diffusion. All the factors/sub-factors and variables have been obtained from the literature and the real world. A close examination of the model reveals that all the factors/sub-factors ultimately lead to sustained use of a KMS. Companies planning to embark on a KMS can consider the variables of the model (see Xu, Quaddus & Wood, 2001) as criteria for the successful diffusion of a KMS. It must be noted that not all the criteria of the comprehensive model will be applicable to all companies. A careful analysis is first needed to select the appropriate criteria for a company. A multiple-criteria modelling approach can then be undertaken to assess the suitability of the company for KMS adoption and diffusion. A similar modelling approach can also be undertaken to find the type of KMS suitable for a company (Quaddus & Achjari, 2001).

As per the factors and variables in the model, an organization can predict whether the system will be suitable for the organization and be accepted by the users when it is embarking on a KMS. The factors and variables in the model will also help the organization to diagnose the possible reasons for unsuccessful implementation and diffusion, and take corrective action to increase the acceptability among the end users and integrate the KMS into its business processes for the purpose of maximizing business performance and optimizing business impact.

Significant Factors of KMS Adoption and Diffusion

This research identifies three of the most significant perception-related variables in the KMS adoption/diffusion process, namely usefulness, user-friendliness and voluntariness of a KMS. The implication is that any KMS has to be extremely user-friendly for any level of users to use it effectively; it has to be useful for the task and a policy must be implemented to facilitate the voluntary use of the KMS. It is noted that any kind of norm (pressure) creation is unlikely to make it grow effectively within the organization.

The results show perceived user-friendliness has more influence on people's use and continued use of a KMS than perceived usefulness, and suggest that an organization should advocate the user-friendliness/ease of use of the KMS when it promotes and implements a KMS. Meanwhile, when the organization is deciding to embark on a KMS, it should pay more attention to the benefits the system can bring to the organization rather than spending too much time on looking for, or designing, a user-friendly system. It should be understood that the ultimate purpose of the KMS is to improve the performance of the organization and individuals. No amount of user-friendliness can substitute for the usefulness/benefits of a system (see Xu 2003).

All the hypotheses related to the sequence of the KMS diffusion process (see Quaddus & Xu, 2003) were found to be significant. This is an important and significant finding. It clearly demonstrates how KMS adoption and diffusion should be planned in Australian organizations. A well-planned sequence must be adopted for the effective adoption and diffusion of KMSs.

End-user Focus and Involving People in the KMS

An organization should build or adopt a KMS, which is user-focused and can bring benefits to everyone in the organization. If individuals do not commit to and support the KMS, it will never happen. Organizations should adopt and implement a user-friendly KMS, otherwise people may not use it even though it is useful. Organizations should only establish a KMS that people want to use. They should shift their attention from the hardware and software perspectives of a KMS, and pay more heed to the effects of people problems and motivational problems (Huber, 2001). If user-centered principles are followed, the required KM activities will automatically be identified as the KM project proceeds (Robertson, 2002). It will be of greater value for everyone in the organization if people shared their knowledge. And everyone can enjoy the benefits of sharing knowledge and using the KMS.

When an organization is embarking on a KMS, it should encourage people's involvement in the system and local ownership of the system, which will have an impact on people's decision to accept and use the system. Users

are both beneficiaries of, and contributors to, the KMS (Alavi & Leidner, 1999); as a result, people must be involved in KMS development and maintenance. In other words, people should be invited to participate in KMS activities all the time and right from the beginning.

At the same time, the organization should ensure that everyone understands that an important part of their working life is to learn as much as possible, contribute knowledge to the system, and participate in the dissemination of knowledge (Pan & Scarbrough, 1998). Also, the organization should motivate and reward early adopters of the KMS. By doing so, the critical mass of KMS use can be achieved, and more people will be willing to share their knowledge and make a contribution to the system (Bansler & Havn, 2002). But the positive incentive scheme aimed at encouraging people's use of, and contribution to, the system should be carefully monitored since an overemphasis on rewarded behaviour may induce people to focus on quantity rather than quality of knowledge submission to the system (Bansler & Havn, 2002).

It is seen that, in the KMS adoption/diffusion process, individual factors/differences of end users and task complexity are the significant factors influencing the perceived usefulness of a KMS (see Quaddus & Xu, 2002). This provides an interesting challenge for would-be adopters of KMSs in Australian organizations. Top-level executives of these organizations should plan carefully as their support does not guarantee a positive influence on the usefulness of a KMS. They must look deeply into the task factors and the end users to see if they are conducive to KMS use. Prior to introducing a KMS into the organization, top management should ensure that: (1) people are involved in the KMS adoption and implementation, (2) people have the experience/skills to use computer systems (both software and hardware), (3) people understand what knowledge the organization has and where and how to locate it, (4) people have time to document what they have done, and are encouraged to try new things. The organization should also examine task complexity by looking at a number of multidisciplinary complex projects, the amount of knowledge required for the business to grow, and the need to effectively track and reapply past best practice.

The mediating effects of perceived usefulness between external variables and KMS adoption/diffusion (see Quaddus & Xu 2002) suggest that organizations probably should put more emphasis on shaping and forming the perceived benefits/usefulness. The organization's effort of implementing a KMS might be more fruitful when management's attention focuses on the development of perceptions of usefulness of the KMS rather than implementation issues. Successful implementation will take place if the formation of perceptions is properly handled (Agarwal & Prasad, 1999).

At the meant time, the significant and positive impact of people's perceptions of voluntary use (i.e., the extent of KMS use to perform the tasks required by management and the extent of KMS use required in their job description) on the organic growth of KMS use (see Quaddus & Xu, 2002) indicates that, management pressure may not make people use the KMS. They have to educate and persuade people to use the system. They have to create an environment in which people feel comfortable and willing to use the system.

Organizational Adjustment to Embrace the KMS

The organization should have a KM strategy, otherwise its KM activities will be somewhat ad hoc. Effective KM practices start with a KM strategy. The KM strategy is the vision for an organization's effort in managing knowledge.

An important factor in ensuring the success of a KMS is to identify the business processes of most value to the organization. An integration of the KMS and business processes can help demonstrate how shared knowledge can generate immediate and highly visible returns to both individuals and the organization as a whole (Mazzie, 2000). The KMS must acknowledge an organization's business and business processes since the ultimate purpose of the system is to bring benefits to the organization and improve its performance.

Organizational culture is the set of values, beliefs, norms, and expectations that are widely held in an organization (Huber, 2001). In order to successfully implement a KMS, an organization should work on building a knowledge-sharing and pro-KM culture as well as an organizational structure. The organization should have a culture in which everyone is prepared to share knowledge and knowledge sharing is rewarded. In other words, an organizational culture in which seeking, sharing, and applying knowledge is the norm in the context of a KMS (Huber, 2001). Changing the existing culture and people's work habits is the first hurdle for most KM programs (McDermott, 1999). Chase (1997), Zyngier (2002), The Conference Board (1997), The Delphi Group (1997), and others state that the lack of a knowledge-sharing culture is the biggest obstacle to KM.

KM involves breaking down barriers. Hence, organizations should not operate in a silo structure. The ideal structure for knowledge sharing and effective KM is the matrix structure and team-based approach, in which the teams work across the organization. The KMS has to comply with the organizational structure. If an organization really practices KM, it should continually make adjustments to the organizational structure to facilitate the process.

Structure also influences the use of a KMS. The way an organization is structured in terms of lines of

responsibilities and accountabilities will reinforce the use of the KMS. Probably the most difficult thing for any organization's implementation of KMS is to set up the structure and build enough definition into it so people can retrieve what they want while allowing flexibility. The organization should encourage the practice of knowledge sharing through the linking of 'communities of practice' (i.e., people who perform similar tasks and functions but who are geographically dispersed), which can result in the organization's use of employees' expertise to maximize efficient product and service delivery (Mazzie, 2000).

Role of Top Management in KMS Implementation

At the same time, top management should take the initiative and express strong support for the KMS by appointing KM positions and allocating a KMS budget. Even though some rigor in implementing the KMS may be helpful, management should focus on persuasion and education.

Top management's role in implementation and diffusion of the KMS lies in creating an environment in which people feel comfortable and are willing to use the system. Although management can provide leadership in KM, the management on its own may not always be sufficient to stimulate and sustain effective KM practices.

In order to make the transferring/sharing of knowledge a common practice in the organization, management has to ensure the implementation of policies and reward systems to encourage knowledge sharing. Management should also lead the knowledge sharing practices by setting a good example themselves. They must become users of the KMS.

Training can enhance the use of a KMS through building up people's competence and confidence to try and adopt it. Training will help end users overcome the fear of complexity of the KMS. Although in-class training can provide people with certain skills and knowledge (explicit knowledge), tacit knowledge is developed through hands on practices.

Technology for a KMS

An integrated, integrative, and adequate technology infrastructure is a key driver for KMSs. KM is not new to organizations. For example, organizations have been practising knowledge coding and transferring, such as employee training and development programs, organizational policies, routines, procedures, reports, and manuals, for many years (Gottschalk, 1999). What is new and exciting in KM is the potential use of information technologies (e.g., internet, intranet, data warehouses, search engines and tools, among many others) to help an organization in managing knowledge in a more systematic and effective way (Gottschalk,

1999; Alavi & Leidner, 1999). In other words, the key elements of a KMS, such as people and expertise, already exist; there is a need of good KM technology to do it in a smart way (Phillips Fox, 1998). Clearly, the current rapid development in technologies has empowered the opportunities for knowledge sharing and combination (Nahapiet & Ghoshal, 1998).

The various KMS technologies and applications can be classified into three categories: database and database management systems, communication and messaging, and browsing and retrieval (Alavi & Leidner, 1999). Some of the common functions of KMS technologies and tools are (1) organizing and sharing/transferring of internal benchmarks/best practices, (2) constructing corporate knowledge directories, such as corporate yellow pages, people information archives, etc., (3) creating knowledge networks and knowledge maps; among many others (Alavi & Leidner, 2001). This study found that the top 10 most widely used KMS technologies were (in order): e-mail and communication systems, internet, databases, intranet, document management systems, customer management systems, video conference, online discussion forum, workflow systems, data warehousing/mining, and search and retrieval tools (see Quaddus & Xu, 2002).

In some organizations IT leads KM, and in others it facilitates KM. The best way is for IT to facilitate KM, and not be KM. Most organizations are not in the IT business and do not exist for IT. Technology only puts different components of the system together. There is no doubt that KMSs need technology since it can provide assistance and facilitation in various dimensions of the KM process, namely: knowledge capture and documentation, knowledge storage, knowledge sharing and access, knowledge reuse and application. Technology can make a KMS happen.

Suggestions for KMS adoption and diffusion

Finally, based on the KMS adoption and diffusion model, which has been proved to be valid and significant, this research offers some suggestions to assist in the implementation of a KMS in organizations:

- q Top-level executives of would-be adopters (organizations) of a KMS should plan carefully as their support does not guarantee a positive influence on the usefulness of the KMS. They must look deeply into the task factors and the end users to see if these factors are conducive to KMS use.
- q The KMS system has to be extremely user-friendly. For any level of user to use it effectively, it has to be useful for the task to be dealt with and a policy must be implemented for the system to be used on a voluntary basis. Any kind of norm (pressure) creation is unlikely to make it grow effectively within the organization.

q A well-planned sequence of KMS adoption and diffusion has been mentioned in Quaddus and Xu (2003), but is reintroduced here since we feel it is essential to the success of KMSs in organizations.

- ⇒ Research the organization's challenges and people's needs regarding knowledge, i.e., identify the important knowledge domains for the organization.
- ⇒ Search for suitable applications for the KMS.
- ⇒ Develop a KMS plan/strategy.
- ⇒ Allocate a budget for the KMS.
- ⇒ Appoint a knowledge manager or a chief knowledge officer.
- ⇒ Build up/set up the KMS.
- ⇒ Test the KMS through pilot implementation on a limited basis in the organization and optimize the system according to feedback.
- ⇒ Work on the organizational culture and structure to facilitate implementation of the KMS.
- ⇒ Persuade and educate people to use the KMS.
- ⇒ Provide people with continuous training and support to encourage their use of the KMS.
- ⇒ Encourage people to go through a process of self-learning.
- ⇒ Implement the KMS throughout the organization.
- ⇒ Cut off people's old means of accessing knowledge.
- ⇒ Develop an organizational-wide interest in using the KMS.
- ⇒ Monitor people's use of the KMS.
- ⇒ Keep providing the knowledge people want in the KMS.
- ⇒ Promote best practice.
- ⇒ Develop and encourage people's sustained use of the KMS.
- ⇒ Make using the KMS a part of the organization's business.

4. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

The results of this research can help organizations that are currently practising KM or are planning to embark on a KMS, by enhancing their understanding of a KMS and providing them with a checklist of the important variables in the KMS adoption and diffusion model with which to do an internal audit to find out how they fare in terms of these variables. The results of this research also provide guidelines on successfully implementing KMSs in organizations. Meanwhile, even though this research was conducted in Australian organizations, its results will apply to different organizations in various countries across the globe because of its generic approach.

An interesting future study could be looking at the differentiation among the types of KMS adopters. According to Rogers' (1995) theory of innovation diffusion, there are five types of adopters. The first to

adopt a KMS are the innovators, who adopt it because of its intrinsic values, including perceived user-friendliness/perceived ease of use. Later, the early adopters adopt it since it is able to provide competitive advantage. Only then the early majority adopt it for pragmatic reasons, such as return-on-investment, cost and benefit. They are followed by the late adopters and conservatives, who wait until the KMS is very well established. Most adopt a new IT for the sake of its extrinsic value, such as perceived usefulness (Moore, 1991). At the mean time, the results of this study were obtained from large Australian organizations. The validity of generalizing these results to other countries or small firms is yet to be determined. Furthermore, this study only reflects and measures a snapshot situation of KMS adoption and diffusion at a particular point in time. Future research could do longitudinal studies to have a better understanding of the process of KMS adoption and diffusion.

REFERENCES

- [1] Alavi, M. & Leidner, D. E., "Knowledge Management Systems: Issues, Challenges, and Benefits", *Communications of the Association for Information System*, Vol.1, Article 7, 1999,.. [Online], Available: <http://cais.isworld.org/articles/1-7/article.html> [2001, May 15].
- [2] Bansler, J.P. & Havn, E., "Exploring the role of networks in IT implementation-The case of knowledge management systems", *The Xth European Conference on Information Systems*, Gdansk, Poland, June 2002.
- [3] Brooking, A., *Corporate Memory: Strategies for Knowledge Management*, International Thomas Business Press, London, UK, 1999
- [4] Gottschalk, P., "Use of IT for knowledge management in law firms", *Journal of Information, Law, and Technology*, 1999, [Online], Available: <http://elj.warwick.ac.uk/jilt/99~3/gottschalk.html> [2002, May 15].
- [5] Huber, G. P., "Transfer of knowledge in knowledge management systems: unexplored issues and suggested studies", *European Journal of Information Systems*, Vol. 10, No. 2, pp. 72-79, 2001
- [6] Mazzie, M., "Key challenges facing the evolution of knowledge management", in *Knowledge Management for the Information Professional*, eds T. K. Srikantaiah and M.E.D. Koenig, Information Today, Inc., Medford, New Jersey, USA, pp. 99-114, 2000.
- [7] McDermott, R., "Why information technology inspired but cannot deliver knowledge management", *California Management Review*, Vol. 4, No. 4, pp. 103-117, 1999.
- [8] Moore, G. A., *Crossing the chasm: Marketing and selling high-tech goods to mainstream customers*, HarperBusiness, New York, USA, 1991.
- [9] Nahapiet, J. & Ghoshal, S., "Social capital, intellectual capital, and the organizational advantage",

- Academy of Management Review*, Vol. 23, No. 2, pp. 242-266, 1998.
- [10] Pan, S. L. & Scarbrough, H., "A Socio-Technical View of Knowledge-Sharing at Buckman Laboratories", *Journal of Knowledge Management*, Vol. 2, No. 1, pp. 55-66, 1998.
- [11] Phillips Fox, *Leading the Way in Knowledge Management*, 1998, [online], Available, <http://www.phillipsfox.com.au/phillipsfox/publications/pbo98006.htm> [2001 July 20].
- [12] Quaddus, M. & Achjari, D., *A Framework for Assessing Electronic Commerce Success: A Blend of Multiple Criteria and Structural Equation Modelling*, Working paper, Graduate School of Business, Curtin University of Technology, Perth, 24 pages, 2001.
- [13] Quaddus, M., Xu, J. and Wood, M., "Factors of Usefulness and Benefits of Knowledge Management Systems: Results of Field and Pilot Studies in Western Australia", *Proceedings of Sixth International Research Conference on Quality, Innovation & Knowledge Management*, Kuala Lumpur, Malaysia, Feb 2002, pp. 1264-1278, 2002.
- [14] Quaddus, M. and Xu, J., "Factors of Adoption and Diffusion of Knowledge Management Systems in Australia: A Structural Equation Modelling Approach", *Proceedings of The Thirteenth Australasian Conference on Information Systems (ACIS 2002)*, Melbourne, Australia, December, 2002, pp. 75-86, 2002.
- [15] Quaddus, M. and Xu, J., 'A Stage Model of Knowledge Management Systems Diffusion Process: Findings From an Australian Study', *Proceedings of 3rd International Conference on Electronic Business*, 9-13 December, Singapore, 2003.
- [16] Robertson, J., *Using Usability to Direct KM Systems*, 2002, [Online], Available: http://www.steptwo.com.au/papers/kmc_usability/pdf/KMC_Usability.pdf [2002, Oct. 12].
- [17] Rogers, E. M., *Diffusion of Innovations (Fourth Edition)*, The Free Press, New York, USA., 1995.
- [18] Sarvary, M., "Knowledge management and competition in the consulting industry", *California Management Review*, Vol. 41, No. 2, pp. 95-107, 1999.
- [19] Sveiby, K., *Small Knowledge Companies-Wave of the Future*, 1995, [Online], Available: <http://www.sveiby.com.au/KnowledgeOrganizationsAust.html> [1999, Sep.16].
- [20] The Conference Board, *Beyond Knowledge Management: New Ways to Work and Learn*, The Conference Board Inc, New York, USA, 1999.
- [21] The Delphi Group, *Delphi Research Identifies Culture as Key Hurdle to Knowledge Management*, 1997, [Online], Available: www.delphigroup.com/pressleases/1997~PR/11051997/KMAApplicationsHurdle.html [1999 Sep. 14].
- [22] Wiig, K.M., "Integrating intellectual capital and knowledge management", *Long Range Planning*, Vol. 30, No. 3, pp. 399-405, 1997.
- [23] Xu, J., Quaddus, M. and Wood, M., "Factors of KMS Diffusion: A Field Study in Australia", *Proceedings of 2nd Europe Conference on Knowledge Management*, IEDC Bled School of Management, Bled, Slovenia, November 2001, pp. 777-793, 2001.
- [24] Xu, J., *A Study of the Factors Influencing Knowledge Management Systems Diffusion in Australia*, PhD Thesis, Graduate School of Business, Curtin University of Technology, Perth, Australia, 2003.
- [25] Zyngier, S., "Knowledge management obstacles in Australia", *Proceedings of The Xth European Conference on Information Systems*, Gdansk, Poland, June 2002.

A Methodology to Support the Design and Deployment of Knowledge Management within Inter-organizational Networks

Tomaso Forzi¹, Meikel Peters², Stefan Bleck¹

¹ E-Business Engineering Department of the Research Institute for Operations Management (FIR) at Aachen University of Technology (RWTH Aachen), Pontdriesch 14/16, 52062 Aachen, Germany

² Chair and Institute of Industrial Engineering and Ergonomics (IAW) at Aachen University of Technology (RWTH Aachen), Bergdriesch 27, 52062 Aachen, Germany

¹ {Tomaso.Forzi, Stefan.Bleck}@fir.rwth-aachen.de, ² M.Peters@iaw.rwth-aachen.de

ABSTRACT

A well-functioning Knowledge Management is a competitive advantage for enterprises that act in co-operative and distributed networks with knowledge intensive production processes. A Knowledge Management approach that integrates both, hard factors (e.g., Information Technology) and soft factors (e.g., cultural aspects) for distributed and dynamic entrepreneurial (inter-organisational) networks is currently missing. This paper presents research findings of a project that is developing a methodology to support a service provider responsible for the KM within distributed entrepreneurial networks.

Keywords: Knowledge Management, Knowledge Networks, Inter-organisational Networked Businesses, Collaborative Networks

1. MOTIVATION AND RESEARCH RATIONALE

In recent years two major trends could be observed in the sector of manufacturing enterprises. On the one hand, the companies concentrate more and more on their core competencies, while on the other hand knowledge is increasingly recognized as a success factor in the tough competition of the global market arena [17] [18].

As a result of the concentration on core competencies, peripheral functions are abandoned and taken over by external partners [11]. This trend leads to a higher productivity and flexibility of the companies' core businesses, but also increases the interaction between different business units [18]. While this gives much more flexibility to the production process as a whole, it also bears the challenge of handling these interactions efficiently.

Knowledge, on the other hand, has been accepted as a crucial factor in business life [13]. Many companies have therefore focused on Knowledge Management (KM) and hence implemented KM solutions [2]. However, such implementations are usually focused on internal, enterprise specific applications and are mainly technology driven [9] [12] [13] [14]. They are therefore suitable only to a very limited extent to support companies in order to face the challenges of KM in distributed business networks.

In addition, different goals, ethic values and cultures as well as a lack of trust inhibit the exploitation of the potentials of networked knowledge [4][6][10]. Appropriate models and methods and in particular a methodology for KM in distributed and globally

dispersed entrepreneurial networks are lacking [6] [16] [19].

Thus, the idea behind our research is that, in order to build up and maintain co-operation and establish knowledge transfer in business networks, the companies involved require external support. Networked enterprises have to be taught how KM along complex and networked value chains can be accomplished. An external knowledge manager who acts as an intermediary in the network and represents a neutral trust centre can analyse applicable tools and critical success factors for KM in networks.

The presented research project integrates both, new Information and Communication Technology (ICT) driven organisational concepts and human-oriented approaches with KM methodologies and instruments, in order to implement an innovative KM Service Provider for distributed networks.

2. A METHODOLOGY TO DESIGN KM WITHIN INTER-ORGANISATIONAL NETWORKS

A methodology to support the design of KM within inter-organisational networks is supposed to fulfil two essential requirements. On the one side it has to describe how KM in the considered network is structured and with which instruments the handling of knowledge can be organised in the network. On the other side, it has to describe how KM has to be implemented in the network, i.e. the methodology has to support the KM-deployment along the complete KM life cycle.

In analogy with the „Aachen PPS Model“ [15], we structured the methodology into a description and a procedure model. In the *description model* all relevant

elements to design and configure KM in business networks as well as their mutual relationships are described. In the *procedure model* it is explained how and in which phases KM can be implemented in the considered network. Within the methodology, also the interrelations between the two models have to be highlighted.

3. DESCRIPTION MODEL

The description model has four elements, which represent the relevant views for the service provider within the design of KM, namely: KM-tasks, Network, Design Areas, KM-Methods and Instruments. As shown in Figure 1, the different views are strictly mutually interrelated.

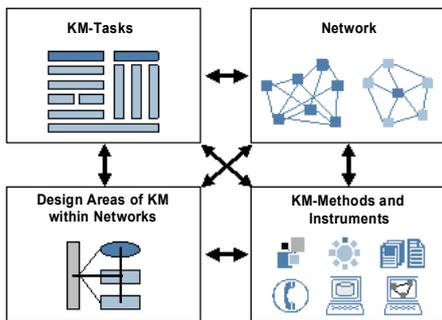


Figure 1: The elements of the description model

3.1 KM-Tasks

The KM-tasks that have to be dealt with are of the most different kinds [4] [6] [14] [16]. We distinguish between direct and indirect KM-tasks.

Direct KM-tasks involve directly the KM of the network; their fulfillment is of great influence on the whole KM process. To the direct KM-task group belong: a) definition of knowledge objectives, b) identification of knowledge, c) acquisition of knowledge, d) development of knowledge, e) distribution of knowledge, f) deployment of knowledge, and g) preservation of knowledge.

On the other hand, *indirect KM-tasks* involve functions that deal indirectly with the KM of the network, but which are nonetheless of great relevance within the conceptual design and implementation of KM within the network. Indirect KM-tasks are: a) KM-controlling, b) management of KM-infrastructure (e.g. IT), c) personnel management, d) fostering / cultivation of knowledge culture and KM-vision, and e) management of knowledge structure.

All such tasks have to be taken into consideration while analysing the existing KM and especially in the following phase, in which the new KM-approaches are modelled and appropriate instruments are selected.

3.2 Network

Distributed entrepreneurial inter-organisational networks can be described according to a variety of characteristics, which have an even wider range of potential expressions [9] [12] [17] [18]. In order to structure the field, a set of clusters of network characteristics was identified: a) *network size and demography* (e.g. number of partners, number of active partners, size of partner enterprises, number of active individuals); b) *legal, contractual and financial characteristics* (e.g. network borders, kind of commitment, funding); c) *economic characteristics* (e.g. position within the value chain of the partners, competitive relations among the partners, industrial sector, network reach); d) *temporal characteristics* (e.g. development phase of the network, time frame of planned cooperation); e) *characteristics related to the network organisation* (e.g. coordination, organisational structure, organisational forms within the network); f) *characteristics related to the management of information and knowledge within the network* (e.g. type of information exchange, type of communication, KM objectives, KM phase).

Clearly, the clusters are open and can be amended with new characteristics and expressions. Based upon all the identified characteristics and related sets of possible specifications a standardized framework for network analysis has been developed.

3.3 Design Areas

The different design areas of KM within inter-organisational networks are presented within the description model of KM within distributed intra-organisational and inter-organisational networks [7] [8]. The four core elements of the Knowledge Management of organisational networks are: Network Knowledge, KM-Processes, KM-Resources and KM-Culture of the network (see also Figure 2).

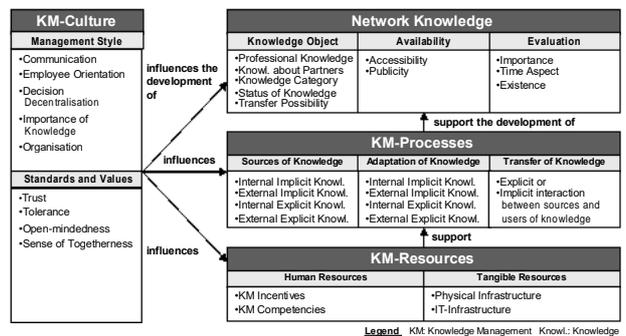


Figure 2: Design areas of KM

Starting point is the actual *knowledge within the network*: Its potential is the reason behind all KM activities. The network knowledge appears in different types of knowledge with specific attributes. It can be retrieved from different sources, and differs according to its accessibility [12] [20].

To realise the potential of Network Knowledge certain *KM processes* are necessary [12] [20]. They can be classified in processes needed to retrieve knowledge, to change knowledge and to pass on knowledge.

Such processes again rely on *appropriate resources* – both human and tangible. For example, employees need certain competencies to be able to carry out KM processes, but they also need the corresponding physical and IT infrastructure to fulfil their tasks [3] [5] [10].

These three areas are all influenced by a fourth one: the *KM culture of the network*. Cultural aspects can enhance an open knowledge transfer or inhibit a positive attitude towards sharing knowledge [1]. Thus a detailed consideration of the management style as well as the standards and values within the network is very important. Different aspects have to be examined in this context: communication, employee orientation, decision decentralisation, importance of knowledge, trust and tolerance etc.

Based upon all the attributes and related sets of possible specifications for each of the four areas an appropriate standardized framework for the analysis of the current KM within distributed Networks has been developed. This analysis framework consists of questionnaires, guidelines for interviews, screen-plays for workshops and tools for each area of the Knowledge Model to help the Knowledge Manager during the analysis of the existing KM in the network. The analysis framework was validated in four networks of industrial partners within the consortium.

3.4 KM-Methods and Instruments

The KM methods and Instruments are of the most different kinds [2] [4] [6] [7] [13] [19]. A vast number of KM-methods and instruments was collected and hence classified according to *a) related KM-Tasks, b) KM-design area of action, c) related form of knowledge representation, d) classification of method typology* (technological, organisational, personnel methods and instruments, see also Figure 3).

IT-Instruments for KM	Organisational Methods and Instruments for KM	Personnel Methods and Instruments for KM
<ul style="list-style-type: none"> • Knowledge Portals • Document Management Systems • Groupware Systems • Data Warehouse Systems • E-Learning-Systems • Information Retrieval Systems • Intelligent Agent Systems • Filter Systems • Workflow Management Systems • ... 	<ul style="list-style-type: none"> • Company Vision • Knowledge Map • Experience Exchange Circle • Competence Network • Experience Documentation • Job Rotation • Job Enrichment • Job Enlargement • Information Mapping • ... 	<ul style="list-style-type: none"> • Coaching/ Mentoring • Network for New Employees • Learning Coach • Interview Techniques • Creativity Techniques • Shadowing • Story Telling • Seminars • ...

Figure 3: Classification of KM-Methods according to their typology

Hence they have been structured into a tool map, which is currently being implemented into a software tool.

4. PROCEDURE MODEL

The procedure model for the deployment of KM within entrepreneurial networks has six phases, which are namely: Initialisation, Analysis, Conceptual Design, Implementation, Operational Management, and Termination.

Figure 4 shows the rough structure of the procedure model for the service deployment. Clearly, the task of the Service Provider for KM within a distributed network accompanies the whole lifecycle of KM within the network.

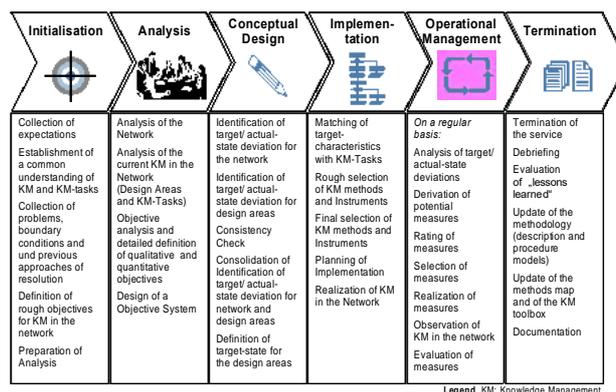


Figure 4: Phases of the procedure model

4.1 Initialisation

In the first phase, the initialisation, the Service Provider has to trigger the whole KM project within the involved inter-organisational network. After the collection of expectations of the different stakeholders, a common understanding of KM and KM-tasks within the considered network has to be established. Hence, current problems, boundary conditions and previous approaches of resolution of the KM issue in the considered network have to be collected, in order to be able to define rough objectives for KM in the network as well as the related evaluation criteria. Eventually, all the collected data and information have to be analysed and structured in order to prepare the following phase of analysis.

4.2 Analysis

In this phase the relevant elements of KM in the network have to be thoroughly analysed, in order to define a concrete basis for the conceptual design of the striven KM approach. In particular, a detailed analysis both of the Network and of the practiced KM in the Network (in particular with the help of the views “Network”, “Design Areas” and “KM-Tasks”) has to be conducted. Thus, within an analysis of KM objectives of the network, a set of detailed qualitative and quantitative objectives have to be defined in order to structure an objective system for the KM of the network. Eventually, appropriate evaluation criteria have to be defined.

4.3 Conceptual Design

Within this phase, the striven KM approach within the network has to be designed conceptually. First of all, the target/ actual-state deviation for the network as well as for the KM design areas has to be identified. Hence, such deviations have to be compared and, if necessary, harmonized in order to assure consistency between the network and the design areas. Eventually, after a consolidation of target/ actual-state deviation for network and design areas, a target state for the design areas can be derived.

4.4 Implementation

Within the implementation phase, with the help of the appropriate KM methods and instruments, the KM has to be implemented within the whole network. After a matching of the target states of the design areas with the related KM-Tasks, an initial rough selection of KM Methods and Instruments with the help of the tool map will be undergone. Thus, from all the methods and instruments potentially suitable for the considered network, a final, more restricted tool set will be selected. With this information the KM implementation will be hence planned and eventually realized.

4.5 Operational Management

The KM has to be then operationally deployed over the whole time the considered network is active. This phase, the operational KM, is the most time-consuming. On a regular basis, following steps have to be conducted within a control loop: basis target/ actual state deviations have to be analysed, potential measures, if needed, have to be consequently derived and rated. Thus, appropriate measures have to be selected and realized. The behaviour of the network has to be then observed over time, in order to be able to eventually evaluate the implemented measures.

4.6 Termination (of Service)

When, for whatever reason, the network terminates to operate or the network management decides not to make use of the offered services anymore, the Service Provider will terminate the service. After the official termination of the service, a series of debriefings with the involved key players will be conducted; hence, the gained „lessons learned“ will be evaluated. If necessary, the methodology (description and procedure models) will be updated, according to the gained inputs. The same applies to a possible update of the methods map and of the KM toolbox. Eventually, a final documentation will conclude the project.

5. INTERRELATION BETWEEN DESCRIPTION AND PROCEDURE MODEL

In order to be able to make use of the methodology, the Service Provider needs a further instrument, that shows which view of the description model (KM-tasks, Network, Design Areas, KM-Methods and Instruments) as well other potentially relevant views (e.g. Objective Model or Evaluation Model) has to be used in a specific phase of the procedure model (Initialisation, Analysis, Conceptual Design, Implementation, Operational Management, and Termination).

Figure 5 shows the basic idea of the interrelation between the two models. Clearly, the extent of such an interrelation has to be specified with the needed detail.

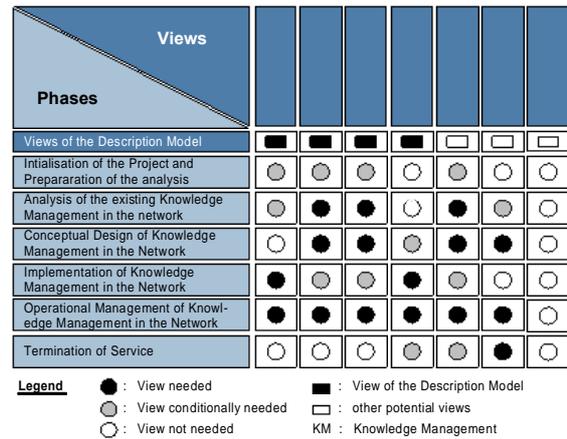


Figure 5: Interrelations between the models

6. CASE STUDY: KM IN AN INTER-ORGANISATIONAL NETWORK

In the following a case study of an inter-organisational network, in which the presented methodology is currently being applied, will be presented. The considered case involves the VIA-Network, a regional inter-organisational network of 20 small and medium enterprises (SME) in the automotive industry located in the German Federal State North Rhine-Westphalia (NRW). In this case study, the role of the service provider is deployed by the VIA-Consult GmbH, one of the network companies, that was founded by other network companies as a consultancy firm for the network.

During the *initialisation phase*, a series of workshops with five selected companies¹ of the VIA-network was conducted. Within this phase, the different expectations of all stakeholders were collected and a common understanding of KM was established. Since, until that moment, no Knowledge Management projects had been

¹ Fischer & Kaufmann GmbH & Co. KG, Kirchhoff Kutsch GmbH, Heinrichs GmbH & Co. KG, Heinrich Huhn GmbH & Co. KG, Krah RWI GmbH. VIA Consult GmbH & Co. KG acted as intermediary between the research institutes and the network companies and carried out the data collection to a large extent.

conducted, no specific problems could be identified in the field.

Hence, in the following phase, a detailed *analysis* of the existing network knowledge was carried out. The objectives of co-operation within the VIA-network are, among others, combined sourcing, know-how-exchange in working groups, joint projects and joint ventures. Since no specific area of cooperation or business processes could be specified for a support by KM beforehand, the main goal of the analysis of the network knowledge was the identification of potentials for knowledge transfer. Furthermore the collected data were also used for the development of a detailed map of the network knowledge in order to assist network partners in finding sources of knowledge for specific topics within the network.

The overall goal of the following phase, the *conceptual design*, was to select appropriate KM approaches for the support of inter-organisational business processes. From the data gained within the analysis, several potential areas for knowledge exchange could be identified. Work tasks that appeared to be suitable for an exchange of knowledge could be found mostly in non-productive areas. In the productive areas, mainly organisational topics were identified. Examples for the identified areas of potential knowledge sharing are: organisational concepts for the flow of goods within the firm, organisational concepts for reducing interfaces within job processing, analysis of customer satisfaction, and establishment of performance figures. As far as work tasks in the productive areas are concerned, the sharing of knowledge used in comparable technical processes should offer advantages. In the study at hand however, the network companies involved appeared to focus on the potentials in non-productive and organisational topics. Several possible explanations for this were identified: on the one hand in spite of technical work tasks, which appear to be similar on the level of abstraction analysed in the study, the companies manufacture specific products, so that they are not directly comparable. On the other hand the companies' core competencies are concentrated in the productive areas. This implies that the exchange of knowledge concerning these areas is more likely to be regarded as critical. A further explanation lies within the fact that the companies' core competencies are considered highly developed so that an exchange of knowledge and experience is sought in other areas.

During the following phase, the *implementation*, based upon the identified potentials for knowledge exchange and the results of the analysis and conceptual design, appropriate KM-measures as well as pilot areas for their implementation were chosen. The objective of such pilots was to initiate and support the transfer of knowledge and to generate best practices for the whole VIA-network. With the help of the method map developed within the project, the following

KM-measures were chosen: thematic working groups, Communities of Practice, as well as Yellow Pages and database-solutions. The areas in which such measures are going to be implemented are: remuneration, emergency planning, treatment of surface, performance figures/ controlling, analysis of customer satisfaction, and total productivity management.

At the current stage, the thematic working groups and the Communities of Practice are being initiated, while the technical solutions for the Yellow Pages and the databases are being selected.

7. CONCLUSIONS AND FURTHER NEED FOR ACTION

The methodology presented in this paper provides a framework for the analysis, conceptual design, implementation and operational management of KM within inter-organisational networks. The methodology consists of a description model, in which all the relevant elements (i.e. KM-Tasks, the network itself, KM Design Areas, KM Methods and Instruments) of KM within networked organisations are included, as well as a procedure model, in which the different phases of the KM life cycle in the network are introduced.

The application of the methodology in the VIA-network was also presented; in particular the phases of initialisation, analysis, conceptual design, and implementation were discussed. The next step will be the operational management over time of the selected KM-measures. In a final step the implemented KM-measures will be evaluated in the VIA-network and hence extended to the other 15 enterprises of the network.

As far as the theoretical framework is concerned, the current research issues are a deeper analysis of the interrelations between the elements of the design areas, since within the case study potentials could be highlighted; eventually, the development of an appropriate set of performance figures to evaluate the effects of KM-measures in the above-mentioned areas is also striven.

ACKNOWLEDGEMENT

The Project „*Der Dienstleistungsmanager im Netzwerk der Zukunft*“ (“*Service Provider for Knowledge Networks*”) is funded by the German Federal Ministry of Education and Research (grant number: 01HW0206). The other partners are: Bauer Maschinen GmbH, the Chair of Product Development at Munich University of Technology, GPS Schuh & Co. GmbH, VIA Consult GmbH, WET Automotive Systems AG.

REFERENCES

- [1] Ashkanasy, N.M., Broadfoot, L.E., Falkus, S. Questionnaire Measures of Organizational Culture. In: Ashkanasy, N.M., Wilderom, C.P.M., Peterson, M.F. (eds.), *Handbook of organizational culture and climate*, 2000, pp. 131-145.
- [2] Bach, V., Vogler, P., Österle, H. (eds.). *Business Knowledge Management: Praxiserfahrungen mit Intranet-basierten Lösungen*, Berlin: Springer, 1999.
- [3] Becker, F.G. Anreizsysteme für Führungskräfte: Möglichkeiten zur strategisch-orientierten Steuerung des Managements. Poeschel, Stuttgart, 1990.
- [4] Davenport, T., Prussak, L. *Working Knowledge*. Boston: HBS Press, 1997.
- [5] Döring-Katerkamp, U., Trojan, J.. Motivation und Wissensmanagement –eine praktische Perspektive. In: Franken, R., Gadatsch, A. (eds.), *Integriertes Knowledge Management: Konzepte, Methoden, Instrumente und Fallbeispiele*, Vieweg, Braunschweig/Wiesbaden, 2002, pp. 133-149.
- [6] Eppler, M.J., Sukowski, O. (eds). *Fallstudien zum Wissensmanagement*. St. Gallen: NetAcademy Press, 2001.
- [7] Forzi, T., Peters, M., Winkelmann, K. A Framework for the Analysis of Knowledge Management within Distributed Value-creating Networks. In: Tochtermann, K., Maurer H. (eds.): *Proceedings of the 4th International Conference on Knowledge Management (I-KNOW '04)*, Graz (A), 30.06-02.07.2004, pp. 432-439
- [8] Forzi, T., Quadt, A., Schieferdecker, R., Stich, V. Service Provider for Knowledge Networks, in Weber/Pawar/Thoben (Editors): *Proceedings of the 9th Intern. Conference on Concurrent Enterprising (ICE 2003)*, Espoo (FIN), June 16th –18th, 2003, pp. 161-168.
- [9] Gebauer, J., Buxmann, P. Assessing the value of interorganizational systems to support business transactions. In: *Proceedings of HICSS-32*, Maui, Hawaii, 1999.
- [10] Grewe, A. Implementierung neuer Anreizsysteme –Grundlagen, Konzept und Gestaltungsempfehlungen. Hampp, Munich, 2000.
- [11] Hagel, III J., Singer, M. Unbundling the Corporation. In: *HBR*, 77(2), 1999, pp. 133-141.
- [12] Klatt, R., Kopp, R. (2004). Wertschöpfungsorientiertes interorganisationales Wissensmanagement. In: *Proceedings of Crosscomp Final Conference*, SFS, Dortmund, 25-26.03.2004.
- [13] KPMG. *Knowledge Management im Kontext von E-Business*. Berlin: KPMG, 2001.
- [14] Krcmar, H.: *Informationsmanagement*. 2. Aufl., Berlin, New York: Springer, 2000.
- [15] Luczak, H., Eversheim, W. (Eds.) *Produktionsplanung und -steuerung. Grundlagen, Gestaltung und Konzepte*. 2. Edition, Springer Verlag, Berlin, 1999.
- [16] Nonaka, I., Takeuchi, H. *The Knowledge-Creating Company*. New York: Oxford University Press, 1995.
- [17] Parolini, C. *The Value Net: A Tool for Competitive Strategy*. NY: John Wiley, 2000.
- [18] Picot A., Reichwald R., Wigand R.T. *Die grenzenlose Unternehmung –Information, Organisation und Management*, 4th edition. Wiesbaden: Gabler Verlag, 2001.
- [19] Probst, G., Raub, S., Romhardt, K. *Wissen managen: Wie Unternehmen ihre wertvollste Ressource optimal nutzen*, 2nd ed., Wiesbaden: Gabler, 1998.
- [20] Schieferdecker, R. Analyse der Wissensarten in Unternehmensnetzwerken. In: Luczak, H. (ed.): *Kooperation und Arbeit in vernetzten Welten*, Stuttgart: Ergonomia Verlag, 2003, pp. 266-270.

A Qualitative Approach to E-Learning ? A Dynamic Situation in Montpellier Business School.

Dominique Drillon¹, Eleni Boursinou²

¹ CEROM - Montpellier Business School, ddrillon@supco-montpellier.fr

² CEROM/University of Montpellier I, boursinou@hotmail.com

ABSTRACT

The article proposes a more productive way and approach for an e-learning research methodology. The e-learning tools: the computer and the Internet are serious obstacles for sense making in the student's behaviour context. Their use defines in a way the user's behaviour. This article underlines the advantages of qualitative methods for this kind of research and vote for an ethnographic method. Our suggestion has many assets in order to understand and consider interactions. Communication processes tend to have a strong influence on learning behaviour and are almost neglected during quantitative methodology. Accurate and corrective feedback is essential for establishing and performing e-learning techniques. Our suggestions are highlighted by lessons learnt from Montpellier Business school experience.

Keywords: E learning, qualitative methodology, context, interaction, business school, France.

This article is an attempt to present the context of a long-term study conducted in Montpellier Business School. We are interested in the e-learning tool and how both teachers and students perceive it so that best results will be potentially attained. We are going to show you the different aspects of the subject by exposing it in terms of a system. A system whose complexity is obvious and thus the difficulties are numerous. The choice of our methodology was made by taking into account this complexity and we have no choice but being post-modern and transdisciplinary. Although results are not yet publishable, many lessons learnt from this experience can be presented. We are in the first phase of our research where context and methodology have an important role in making choices for the project in the future. In a global forum where both researchers and practitioners come together and share ideas as defined in the goal of this congress, our ambition is to share these ideas and exchange point of views and lessons learnt in the same field from colleagues. Theoretical significance is an important issue in academia, where construct validity, internal validity, generalizability and reliability of the research finding are hallmarks of good quality research. (e.g. Tahai and Meyer, 1999; Scandura and Williams, 2000)

1. THE CONTEXT

Montpellier Business School is doted with an e-learning tool since 2001. The first initiative was made in order to help students that during an exchange one-year-program with affiliated business schools all over the world could "attend" the same course with the ones in the Montpellier campus. As the years went by, e learning was introduced to all students regardless of their place of stay as a new learning tool. Both students and teachers were reluctant to this change. We'll see how we overcame this resistance to change progressively by measures taken for motivating both students and teachers to use this new tool.

"Our conceptual point of departure is that activities of individuals become meaningful only if they are studied in a systemic context, within the comprehensive social system. Research inspired by the qualitative paradigm strives at doing research in, rather than on organizations and has specially made use of the case study methodology." (Rouse and Daellenbach, 2000; Mintzberg, 1999; Eisenhardt, 1989).

Goffman as a product of the Chicago School emphasizes a qualitative analysis of the component parts of the interactive process. He explores the details of individual identity, group relations, the impact of environment, and the movement and interactive meaning of information. Interaction is viewed as a "performance" shaped by environment and audience, constructed to provide others with "impressions" that are consonant with the desired goals of the actor. It is not the interface that counts but the interaction. And most specifically we are interested in situated interactions/artefacts (in situ) and understand the users during these situations.

A learning situation is quite complex to comprehend. Jean Piaget is one of the most important theorists in the field of learning. According his studies, the central mechanism for the child to develop his intelligence depends on two complementary operations: assimilation and adaptation. The first consists in integrating new data in mental structures that already exist. The second is about changing the established structures and adapting to more complex structures. The young child explores his universe by tools he disposes such as his hands, his mouth and all five senses. We have observed the same mechanism on young students during class. As observed by René Spitz this type of mental and cognitive activity necessary to our personal development can be blocked by affective perturbations.

Another mechanism reinforces learning: the transfer between student and teacher.

It is not about the knowledge transfer but the affective “charges” of the student projected to his teacher. It is a mechanism that reminds us the psychoanalytic transfer. In all relationships this transfer exists even if actors are conscious or not.

This mechanism can be compared to what is commonly known as the Pygmalion effect. Robert Rosenthal studied this mechanism in the 60’s at the University of Harvard. He concluded that if the teacher believes in students’ chances for succeeding he could positively influence their performances. If on the other hand he has doubts, students are very likely going to confirm his belief. How is this Pygmalion effect going to be transposed in an e-learning situation?

We chose not to study cognitive activities such as memorization and motivation and their impact on learning. It is obvious that they play a key role. We should underline the paradox between the necessity to adapt and at the same time the resistance to change for seeking security. Man is between two opposed urges: life and death, Eros and Thanatos.

The e learning has a shattering effect on our way of learning and teaching as well. We are continuously seeking to adapt to a new environment. But what type of change is at stake? By putting aside financial stakes that will not be studied in this paper, we will be mostly interested in a comprehensive approach to established practices. What is the future of teachers? What are we expecting from them? Should they be the course “designers”, the tutors or both? What will this change bring about to their every day habits?

The massive use of information and communication technologies, as the Internet and the pedagogy used to accompany this tool ought to be innovative. The teacher is looking for means to achieve the best results and establish a particular methodology. He is constantly searching tools for his pedagogical activity. The innovative technology, the teacher’s distance and the differences compared to a traditional class impose new findings to manage this new learning situation.

Teachers must become information architects. Elaine and Roger Everden have focused their work in this area defining their branch of information architecture more generally:

“Information architecture is the foundation for managing information in general as a corporate resource. It describes the theory, principles, guidelines, standards, conventions and dimensions that are necessary to design an effective management framework for information. Its purpose is to design information structures that help people to use information in effective, productive and innovative ways.”(Everden and Everden, 2003)

A community of teachers is born where cooperation and solidarity are omnipresent not only between teachers but also between them and students.

Each student creates his own “strategy plan” for achieving the goal of the course. He is his own manager of time (rhythm) and pedagogy. They have the total freedom to proceed as they wish. Students are encouraged to evaluate themselves by quizzes (self-evaluation). The simplified teacher-transmitter, student-receiver sketch was replaced by the system teacher-student where the first is not the only knowledge holder but also a mediator.

2.CHANGE AS CO-EVOLUTION

“Resistance to change is considered as an obstacle to change implementation. People resist change because they feel uncertainty about their future, are concerned with potential loss of territory or are afraid to fail in front of new expectations and tasks.”(Wanberg and Bans, 2000)

Some key elements to successful change management are:

- “Assessing organizational culture with an emphasis on informal and formal communication and structure
- Supporting participation and involvement
- Understanding people’s fears and concerns and
- Setting up a training strategy aiming at offering education and support.

These tactics all take into account the importance of the human factor as a critical success or failure factor for understanding and managing change programmes.”(Vakola, 2002)

Technology is a key enabler for implementing change. Students already familiar with the Internet, game boy and other videogames shouldn’t have many problems to adapt. A study conducted by Sherry Turkle shows that the reaction in front of the computer varies and is very personal. French psychiatrists have used computers in a therapeutical context and can enumerate many advantages.

The computer is the mediator, the artefact that facilitates interpersonal exchange.

“Change is a never ending process of continuous transformation, as it ought to be in a world where Darwin’s motto about the responsiveness to change seems more “updated” than ever: It is either the strongest species that will survive, nor the most intelligent, but the ones most responsive to change”. (Prastacos, 2002)

The direction of Montpellier Business School taking into account the importance of the project is establishing a new organizational structure in order to motivate all the

personnel to take part in it. Teachers need to attain a certain level of confidence on this tool as Watzlawick has underlined in order to obtain change. The hierarchy of the institution is doing everything for helping them to obtain this confidence. Advantages for teachers included less class hours, training and technical support, more research activities and free Internet access.

3.METHODOLOGY CHOICES

“Despite the so-called interpretive or hermeneutic turn in the social sciences that has also influenced and, some would argue, undermined this discipline (Reinhardt, 2002), ethnography has clearly established itself as an important alternative to other qualitative approaches as a method for the applied social sciences.”

The definition of ethnography by Hamersley & Atkinson (1995) is one of the most quoted in the field: “... we shall interpret the term “ethnography” in a liberal way, not worrying much about what does or does not count as examples of it. We see the term as referring primarily to a particular method or sets of methods. In its most characteristic form it involves the ethnographer participating, overtly or covertly, in people's lives for an extended period of time, watching what happens; listening to what is said, asking questions—in fact, collecting whatever data is available to throw light on the issues that are the focus of the research”.

According to Garfinkel (1991), “there are good reasons for ethno methodological studies to specify the production and accountability of immortal, ordinary society—that miracle of familiar organisational things—as the local production and natural, reflexive accountability of the phenomena of order. Among those reasons is making discoverable one of those phenomena of order, but only one, namely what analysis incarnate in and as ordinary society, as practical actions locally and interactionally produced and witnessed embodied details, could adequately be”.

Observational methods by which an individual or individuals gather data on processes and behaviours are being chosen. They provide us with the opportunity to collect data on a wide range of behaviours, to capture a great variety of interactions and to openly explore the evaluation topic. We can develop a holistic perspective, an understanding of the context within which the projects operates. This may be especially important where it is not the event that is of interest but rather how the event fit into or is impacted by a sequence of events. Observational approaches also allow the evaluator to learn about things the participants or staff may be unaware of or that they are unwilling or unable to discuss in an interview or focus group.

The most fundamental distinction between various observational strategies concerns the extent to which the observer will be a participant in the setting being studied.

“The ideal is to negotiate and adopt that degree of participation that will yield the most meaningful data about the program given the characteristics of the participants, the nature of staff participant interactions, and the socio-political extent of the program”. (Patton, 1990)

As Hamersley suggests: “to rely on what people say about what they believe and do, without also observing what they do, is to neglect the complex relationship between attitudes and behaviour.”

It is not uncommon to think of organisations as systems (Parsons, 1951; Luhmann, 1971). More precisely: “a system is a structure of interacting, intercommunicating components that, as a group, act or operate individually and jointly to achieve a common goal through the concerted activity of the individual parts”(Thomas 1979). The study of ethnography is automatically the study of some form of organisation or system.

The relationship between people, events and activities is not context free. It can only be understood fully by an appreciation of the context and way in which people interpret their world. Starting by questioning how we make sense of our world in terms of our values, attitudes, actions and interpretations is an essential first step. Theories about what should be done and ways of acting are themselves the product of previously “taken for granted” action. Questioning assumptions and unravelling taken for granted views and perspectives helps to set new co-ordinates and thereby re- defines problems and issues (Aguinis, 1993; Schein 1992; 1996).

Orlando Fals Borda had written about action research: “action research as investigating reality in order to transform it”. “Action Science seeks to change work based and interpersonal behaviour and as such is particularly relevant for business organisations (Argyris, 1999; Argyris & Schon, 1991; Putman, 1999). It is an inquiry approach suited to an organisational culture, which appreciates double-loop learning and is committed to interpreting knowledge in a way that reveals organisational patterns, processes and defensive routines. In Action Science, intervention is psychological since it explores innermost feelings and emotional reactions, some of which are protected by personal defence mechanisms. As these defence mechanisms break down individuals may feel vulnerable and exposed. Risks to self and others involved are reduced if the group themselves are sensitive to others feelings and ensure participants finish sessions on a positive note.”

Dialogue between researchers and teachers and students was conducted during the three years in form of interviews. The research report is said to be a joint construction as a result of these dialogues. Each student is a different case study and all we can conclude is some trends. “Case to case generalization is based on

analogical reasoning. Such reasoning is only plausible when there are solid arguments that, when a particular researched case has characteristics which are relevant for the research conclusions, another case that has not been also researched also has these characteristics.”(Miller, Fredericks, 2003) How qualitative data become evidence for a claim? By blending observational methods for studying dynamic phenomena that change over time.

In order to comprehend an e-learning situation, we will conduct interviews, we'll make a case study that will be in situ and also take part in observations. We are in the first stage where about 20 interviews were made with teachers involved and reluctant ones as well as students. The sample of students was considering a large variety of behaviours: the enthusiastic, the indifferent, foreign students as well as potential students not familiarised with the tool yet

4.CONCLUSION

Rapidity, adaptability, flexibility and autonomy are some of the key qualifications in the beginning of the third millennium. A good manager ought to be “intelligent”and proactive. In this context, students in business schools must learn to manage “change”and being able to adapt in different situations

It is quite delicate to make any kind of conclusions in this stage. It is more about asking a series of questions. If this questioning is pertinent our ambition is fulfilled.

Frontiers are abolished and distance is a key term we should not continue ignoring. E-learning education not only offers us the possibility to learn more and differently but also familiarizes us with the notion of distance that is no more an obstacle but a “banal” element of every day life.

BIBLIOGRAPHY

- [1]Barou G. L., Brouillard E., *L'informatique et ses usagers dans l'éducation*, PUF, Paris, 1996
- [2]Bellier S., *Le e-learning*, Editions liaisons, Paris, 2001
- [3]Boursinou E., Zalonis D., *Le gestionnaire rencontre l'anthropologue : effets de synergie pour un projet d'intervention thérapeutique*, Crossing Frontiers in Quantitative and Qualitative Research Methods, Lyon, France, March 18-20, 2004
- [4]Denzin N., Lincoln Y., *Handbook of Qualitative Research*, Sage Publications, 2000
- [5]Goffman E., *The presentation of self in everyday life*, Garden City, New York, 1959
- [6]Laguerre C., *Ecole, informatique et nouveaux comportements*, L'Harmattan, Paris, 1999
- [7]Miller, S., Fredericks, M. *The nature of “evidence” in qualitative research methods*, International Journal of Qualitative Methods, article 4, 2003
- [8]Perriault J., *La communication du savoir à distance* L'Harmattan, Paris, 1996
- [9]Piaget J., *De la pédagogie*, Odile Jacob, Paris, 1998
- [10]Prastacos G.P., Söderquist K.E., Spanos Y.E., Wassenhove V., *An integrated framework for managing change in the new competitive landscape*, European Management Journal, vol. 20, n°1, 2002
- [11]Probst G. Gibbert M., *Co-operative case writing: A new approach for reconciling the debate between theoretical significance and practical relevance in management research*, Die Unternehmung, Paul Haupt, Bern, 2002
- [12]Smaling A., *The argumentative qualitative research report*, International Journal of Qualitative methods, 1 (3), Article 4, retrieved from <http://www.ualberta.ca/~ijqm>, 2002
- [13]Turkle S., *Les enfants de l'ordinateur*, Denoël, 1986
- [14]Vakola M., *Exploring resistance to change*, Innknow News, n°1, 2002
- [15]Watzlawick P., Helmick J., Jackson D., *Une logique de la communication*, Seuil, Paris, 1967

Adaptive User Interfaces for Intelligent E-Learning: Issues and Trends

Abdul-Rahim Ahmad¹, Otman Basir¹, Khaled Hassanein²

¹ Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

² DeGroote School of Business, McMaster University, Hamilton, ON, Canada
arahim@uwaterloo.ca; obasir@uwaterloo.ca; hassank@mcmaster.ca

ABSTRACT

Adaptive User Interfaces have a long history rooted in the emergence of such eminent technologies as Artificial Intelligence, Soft Computing, Graphical User Interface, JAVA, Internet, and Mobile Services. More specifically, the advent and advancement of the Web and Mobile Learning Services has brought forward adaptivity as an immensely important issue for both efficacy and acceptability of such services. The success of such a learning process depends on the intelligent context-oriented presentation of the domain knowledge and its adaptivity in terms of complexity and granularity consistent to the learner's cognitive level/progress. Researchers have always deemed adaptive user interfaces as a promising solution in this regard. However, the richness in the human behavior, technological opportunities, and contextual nature of information offers daunting challenges. These require creativity, cross-domain synergy, cross-cultural and cross-demographic understanding, and an adequate representation of mission and conception of the task. This paper provides a review of state-of-the-art in adaptive user interface research in Intelligent Multimedia Educational Systems and related areas with an emphasis on core issues and future directions.

Keywords: Adaptivity, Adaptive User Interface, E-Learning, Preference Modeling, User Modeling, Multimedia, HCI

1. INTRODUCTION

The emergence of such revolutionary technologies as Graphical User Interface (GUI), Artificial Intelligence (AI), Soft Computing, Multimedia, Hypermedia, Internet, etc. has created an intense interest in computerized and ubiquitous educational systems like Intelligent Multimedia Educational Systems (IMES) with adaptivity at its core [12], [21], [28]. Indeed, individualized teaching is the most favored practice of instruction and different teaching strategies work best in different contexts. Consequently, the ability to adapt to an individual learner's needs as well as context could significantly stimulate both the learning process and the user engagement [13], [21], [26]. As such, an IMES can contribute to the success of learning if it adequately represents the domain tasks, concepts, and learning goals and intelligently adapt its presentation in terms of complexity and granularity according to the cognitive level/progress of the learner [21], [27].

The development of such pervasive enablers as the Web and Mobile Technologies has further highlighted the research in Adaptive User Interfaces (AUI) offering tremendous educational opportunities by removing temporal and spatial constraints as well as providing personalization and interactivity [10], [12]. However, it is the Web and Mobile Learning Services where the problem of 'being lost in hyperspace' becomes especially critical [13]. As such, the 'information overload' is no more a trite buzzword but a frequently encountered reality for many. Consequently, an adequate educational framework for IMES should focus on cognitive development and knowledge acquisition, through creative, efficient, efficacious, and intelligent tutoring strategies for presentation of the domain knowledge. An AUI in IMES is geared towards seeking

superior outcomes in the spatial and temporal separation as well as presentation of domain knowledge based on user's explicit and implicit preferences, constraints, cognitive progress, decision styles, learning objectives, and access modes.

The AUI has been an object of attention since long and has a significant role in IMES from cognitive, pedagogical, psychological, and social aspects. Adaptivity in IMES could come in terms of adaptation of multimedia content, presentation, navigational options, teaching strategies, etc. to user's needs [45]. In addition, the dynamics of today's society underscore the need for adaptive interfaces accessible by different user groups, including disabled and elderly [1]. However, the potentially diverse scope and heterogeneity of the target population pose greater challenges [27].

In view of diversity, subjectivity, and tedium involved in pertinent concepts and issues, we provide only a brief overview of such concepts. Furthermore, a few capabilities and caveats in the design and deployment of such systems are highlighted. Various promising and synergistic technological options for design and implementation of an AUI are presented for reference purposes, as well.

2. INTELLIGENT INTERFACE PARADIGMS

Research in intelligent user interfaces has been built on such notions as natural language understanding, explanation systems, intelligent tutoring systems, intelligent help systems, computer-aided instruction, multi-modal systems, model-based development, intelligent presentation systems, and agent-based interaction. As a result, several paradigms have been established. The most prevalent paradigms include

user-based interactions, model-based interaction and agent-based interaction [5], [16]. An IUI may facilitate the learner's selection through either *adaptive ordering* of options (based on frequency of use data), *adaptive prompting* (based on contextual information resulting from a model), or *guidance* (with the help of an agent). The boundaries of different paradigms and technologies are quite fuzzy. Nevertheless, here we provide a brief overview of the most popular paradigms.

2.1 User-based Interaction

One frequently cited indication of intelligence in UIs is the ability to adapt the output to the level of understanding and interests of individual users, typically served by adaptivity, adaptability, or dynamicity of the aspects of interaction [34]. An *adaptive* UI changes dynamically in response to its experience with learners [24]. Such an effect is typically achieved by using dedicated software tools responsible for acquiring and maintaining user models and reasoning towards suitable interface adaptations. In contrast, an *adaptable* UI provides tools that allow the user to tailor certain aspects of the interaction in using a system [13], [34]. Such systems allow the user to modify certain aspects of the interactive behavior. Whereas, in a *dynamic* UI the user's behavior is monitored just like with *adaptive* UI. However, instead of changing (adapting) a predefined presentation, dynamic hypermedia systems generate a presentation from "atomic" information items. The research in user-based interaction design is built upon advances in intelligent tutoring systems, learner modeling, explanation systems, and knowledge representation.

2.2 Model-based Interaction

Model-based Interaction development involves the use and articulation of reusable models and knowledge repositories encapsulating the wide variety of details pertaining to the UI development. It promises to decrease both the time and expertise required to create interfaces, through reusable models, automation, decision, and design support mechanisms. It enables design support such as critiquing, design refinements, and incremental updates. However, the current generation of model-based interface development tools has not appropriated the full range of the aforementioned potentialities [5].

2.3 Agent-based Interaction

This paradigm involves the use of software agents to delegate tasks concerning various facets of interaction [5], [15]. Although there is no agreed upon definition of intelligent agents, most intelligent agents are characterized by autonomy, adaptivity, pro-activity, and sociability [15], [31]. They usually contain a representation/model of belief in the state of the environment and have facilities to discover from

patterns of behaviour from user(s) and agent(s). There are three major types of agents involved in agent-based interaction: *interface agents* tie closely to an individual's goals, *task/tutor agents* involve processes associated with arbitrary problem-solving tasks, and *information/domain agents* connect to source(s) of information. An interface agent learns from user's actions working as an intelligent assistant. Whereas, tutor agent provides scaffolding (or envelope) to the learner and progressively removing it as the learner internalizes the knowledge. As such, these agents also require models of the learner, tutor, as well as domain in order to perform their tasks.

3. ADAPTATION TECHNIQUES

As already mentioned, adaptation in a UI can be served by adaptivity, adaptability, or dynamicity in the aspects of interaction. In IMES perspective, adaptation to both user and context are needed. Adaptation to user/learner requires a learner model containing attributes of the learner to which adaptations are sought. At the same time, adaptation to context requires such knowledge models as teacher model, domain model, and interaction model. The adaptation effect may be realized through adaptivity in content selection, presentation, recommendation, etc. [11]. Adaptation may also furnish navigational support [13]. Indeed, different adaptation techniques work most efficiently in different context and require meticulous selection of the most relevant technique as well as the need for adaptation in the adaptation technique(s), or meta-adaptation, during the cognitive progress of the learner [6], [13].

In IMES context, the adaptive navigation support is one of the simplest, earliest, and most extensively studied aspects of automatic adaptation. The underlying notion is to help users find their paths in hyperspace by adapting link presentation to the characteristics/model of the user. The popular means for incorporating adaptive navigation support are direct guidance, sorting, hiding, annotation, and generation of link structure suitable to the learner characteristics [13]. These techniques are used to achieve both global and local guidance/orientation support. Studies have demonstrated the effectiveness of various adaptive navigation support techniques in terms of browsing and learning efficiency [13]. Adaptive content selection is another aspect of adaptivity resulting from the varying interests and preferences of various users regarding available information. It allows information pieces relevant to learner's cognitive level and/or goals to be presented to the user. However, the selected content must be presented in a form appropriate to the user. Consequently, a natural extension to this would be the adaptive content presentation that exhibits different visual layout and/or media. Adaptive recommendations also form an important dimension in achieving effectiveness of an AUI by providing users suggestions about the future line of action.

An ideal scenario would be where the adaptive system has all or most of the adaptation techniques at its disposal. However, it requires a sound understanding of applicability and limitations of each adaptation technique as well as their combinations. Currently, there is a relative dearth of such complex and extensive studies because of obvious reasons.

4. CORE ISSUES IN AUI

Significant research has been done in developing AUIs for accommodating the heterogeneity and evolution of user characteristics. It draws from such diverse fields as psychology, cognitive science, ergonomics, human-computer interaction, AI, etc. However, here we focus on some core issues in AUI within the IMES context.

The goal of adaptivity in IMES is to promote efficient learning rather than simply accommodation of a learner's preferences highlighting the need for a sound understanding of both user and context. The general requirements for a system adaptive to the user include an underlying theory associating user behavior to interface needs, an access to behavioral cues, a variety of interface designs alternatives, and knowledge models accumulating behavioral cues and needs [34], [21], [39]. Nevertheless, the need to make adaptivity actually work in commercially deployable systems while meeting the usability and acceptability requirements poses notable constraints on the theory [34].

A universal problem in any hypermedia navigation is cognitive overload and disorientation [13]. However, the keyhole effect resulting from the contextual, spatial, and temporal separation of information in IMES provides an opportunity to reduce the learner's cognitive overload and distraction [27]. For instance, the amount of information and its context may be constrained in early stages of training with progression based on user's current cognitive and learning capabilities etc [27].

Some possible negative effects of AUI are related to usability, privacy, and trust issues that should be adequately addressed [12], [29]. Indeed, the very idea of adaptivity violates several well-defined and accepted usability principles largely developed for rigid direct manipulation interfaces [41], [42]. Moreover, frequent adaptations reduce consistency in the UI, a much sought for goal in HCI, hampering the learning rate [6], [42].

For instance, a sudden and automatic change in the interface may confuse, disrupt, and frustrate the user. Moreover, the notorious Production Paradox suggests that learners may not adopt/learn strategies that improve long-term efficiency and efficacy. Instead, a learner may adopt a strategy that helps in accomplishing desired task on an ad hoc basis and adapting to such 'quick and dirty' strategies could even prove counter-productive. Absolute, or near absolute, user control is another important usability consideration that is difficult to achieve in AUIs. It has been recommended that user

must get a 'sense of control' over automatic adaptations by making the system 'scrutable' [17]. However, studies in adaptive E-stores have indicated that, when adaptivity is based on implicit information, users want control of both the content and the context [6].

In contrast, AUIs serve many usability guidelines better than rigid interfaces such as reduced information and cognitive load, enhanced task support and visibility of relevant objects or actions, etc. Some researchers have even suggested that an AUI can be effective primarily when actions available to the user remain same.

Like any intelligent system, an AUI may make mistakes in determining the implicit intentions of the learner. The time needed to learn any user-controlled recovery mechanism hampers the system acceptance. Even the simplest of such recovery approaches may result in an increased cognitive load, confusion, and distraction of the user from the intended task. In addition, such user-controlled recovery mechanisms reduce the pedagogical efficacy of the system.

The foremost task in IEMS is to define such design goals as generalizability, scalability, portability, central data storage and ubiquitous access as well as higher responsiveness, learning rate, user engagement, user satisfaction, and reusability, etc. [15], [17], [10]. Nevertheless, the complexity, diversity, volatility, subjectivity, and multiplicity of considerations in designing an AUI mean that our expectations from such systems should be realistically limited.

The next important task in an AUI is the creation of knowledge model, as discussed later. Nevertheless, the richness of the human behavior, technological opportunities, contextual information, as well as multi-sensory nature of the IMES mean there is no panacea in such sensitive human-computer interaction area. Conceivably, the majority of adaptive systems in IMES context are research-level systems [12], [13]. Currently, Web courses present same static learning material to students with diverse cognitive and contextual goals, preferences, and capabilities [13]. Nevertheless, user-modeling experiments have confirmed the importance of a sound theoretical foundation operating in synergy of such practical issues as performance, reliability, and usability.

In IMES context, modeling the pedagogy and linking it to various user characteristics or learning styles is an important determinant of superior learning. Consequently, it is desirable to use the pedagogical strategy most suitable to the learner. As such, the pedagogy would also play an important role in determining the granularity and style of presentation.

5. KNOWLEDGE MODELS IN AUI

An AUI can adapt to either the user or the context of

user's work. However, an effective AUI requires an ability to adapt to both user and context. This requires extending the basis of adaptation by complementing the classic user models with models of context such as purpose, subject domain, pedagogy, interaction mode (platform, location, time, bandwidth) etc. [13]. It means learning theories, concepts, pedagogies, and their impact on the instruction design and practice are important [27]. Consequently, various knowledge models are required in achieving adaptivity in IMES.

Primary models include user/learner model, domain expertise model, pedagogical/tutor expertise model, and interface model. A Learner Model (LM) captures the knowledge about the user for the system to respond to the needs of the user efficiently. A Domain Model (DM) represents the features of the particular domain that is of interest to the user. A Tutor Model (TM) holds the knowledge, capabilities, assumptions and limitations of the tutoring system itself. An Interface or Interaction Model (IM) possesses the dynamic representation of the dialogue between the user and the system. The boundaries of these models can be quite fuzzy and some other modules/models may also appear. Nevertheless, the separation of contents, instructional philosophies, and adaptation options or concept structures facilitates conceptualization as well as system maintenance.

5.1 Learner Model

The main objective of an AUI in IMES is to tailor the learner's information space by presenting learning material according to learner's cognitive level/progress, socio-cultural attributes, goals, plans, tasks, preferences, and beliefs [15]. This objective is highly dependent on the maintenance of adequate, efficient, and reliable LMs by explicitly representing and updating the user's characteristics, preferences, and behavior. This information is used by other components of the system for providing adaptivity. It has been frequently argued that an LM can result in improved interaction by removing the dissonance between learner's cognitive abilities and demands of interactions in IMES [8].

An LM tends to address three general purposes: inferring the user knowledge or general ability (Knowledge Assessment), recognizing the user plans or goals (Plan Recognition), and predicting the user inferences and future behavior (Action Prediction). It should be capable of representing a learner's multiple interests and misconceptions as well as flexible enough to adapt to changes in a user's cognitive level due to interaction with information. Consequently, an LM must deal with the uncertainty in making inferences about a user under incomplete and uncertain information, raising important issues discussed later. Nevertheless, studies have shown that adaptive student modeling can be remarkably robust and contribute to improved learning in IEMS environment [16], [34].

The initial LM could be generated by default values, a set of stereotype-based LMs, or querying learners. However, such stereotypical models lack coarse granularity. Furthermore, querying the users to build initial model increases cognitive and ergonomic load on users. Consequently, an effective LM needs to be constructed unobtrusively based on user behavior using both long-term and short-term information.

5.2 Domain Model

The Domain Model (DM) is the abstract representations of the target subject area. It deals with the link relationships between the concepts and the decomposition of concepts in a structured hierarchy of sub-concepts and atomic information such as texts, images, sounds, and videos [33]. Such conceptual representations should enable the system to analyze, understand, explain, communicate, or predict some aspects of the domain. Tasks, objects, and data form basic building blocks in various DM paradigms. The Task based approaches seek to represent the domain in terms of tasks. The Object-oriented paradigm represents the domain in terms of the objects, the relationships between objects, and exchange of information between objects. Data-Oriented approaches tend to abstract the structure of the domain as a network of entities and the functions of the domain by a network of dataflows and processes.

A DM mimics the expert system paradigm with an ability to generate multiple knowledge-based superior solutions instead of one ideal outcome, thereby endowing flexibility and robustness [3]. Such an approach would involve elaboration, articulation, enumeration, and ranking of competing design alternatives through propagation of design knowledge into the development cycle and embedding design recommendations into the interface implementation [5]

5.3 Tutoring Model

The Tutoring Model (TM) provides a model for the pedagogical philosophy allowing adaptation to different learning styles [37]. It should be able to dynamically incorporate individual differences among students. As such, it requires substantial efforts in knowledge acquisition and representation. TMs are primarily constructed around defined problem solving tasks and cognitive models, so they can interpret each student action (*model tracing*) and draw inferences about the student's knowledge state from each student action (*knowledge tracing*) such that the pedagogical decisions reflect the needs to each student [37].

5.4 Interface Model

The interface is the communication between the student and the aspects of the system. As such, an Interface Model (IM) is required for intelligently controlling the

dialogue, the screen layouts, and other interactions with the learner. Each interface element is linked to an internal representation of an instructional goal and the current state [16]. The two main aspects considered in IM are multimedia content and user exploration. The use of multimedia objects in IMES promises instructional efficacy. However, the complexity of the learning activity involving such a variety of tasks as information retrieval, navigation, and memorization means that just a naïve collection of multimedia objects does not promise improved, or even proper, learning [37]. Various tasks in AUI have different requirements and consequently need different multimedia objects for efficient and efficacious interaction. Consequently, learners need to be provided with diverse forms of interactivity that match the pedagogical objectives of the IMES.

6. UNCERTAINTY MANAGEMENT

In any human modeling, making inferences about the beliefs, abilities, motives, and future actions of people require a good deal of uncertainty management that may manifest as incomplete, inconsistent, imprecise, or uncertain information [4], [8]. Incompleteness suggests the unavailability of some of the information.

Inconsistency refers to the difference or conflict in the knowledge elicited from implicit or explicit information. Imprecision refers to values that are vaguely defined or measured inaccurately. Uncertain information points to the subjectivity in estimate about the value/rule.

The information available for inferencing in AUIs is inherently imprecise, vague, ambiguous, and incomplete with greater gap between the available evidence and drawn conclusions using fairly meager and haphazardly collected data [46]. The knowledge-intensive nature of various models in AUI implies that a little uncertainty in information may translate mischievously into the system response [47]. Consequently, any AUI decision mechanism requires robust ways of coping with such uncertainties. Unfortunately, the majority of theories and tools devised to handle subjectivities and uncertainties in information are quantitative in nature relying on crisp data. Such tools, in general, cannot handle the subjectivity and uncertainty emanating from all aforementioned sources [3], [35]. Thus, formulating effective ways of analysis and revision of knowledge models in AUI under uncertainty is an important research direction. Early solutions to this problem were based on heuristics or ad hoc techniques. Other approaches opted to constrain learners follow a predetermined line of reasoning, making it explicit. Nonetheless, such approaches could result in inflexible and overly constrained interface.

However, certain numerical uncertainty management tools have gained prominence. These methods for tackling uncertain knowledge are generally referred to as Soft Computing (SC). The role model for SC is the human mind and differs from conventional or hard

computing in its tolerance of imprecision, uncertainty and partial truth. As such these hold immense promise for AUI.

6.1 Deterministic Approaches

Deterministic approaches work under the simplifying assumption that all the required information required can be quantified a priori and made available in need [3]. Such approaches usually make use of arbitrary default, user-defined, or expected values that are possibly refined by the user during the course of interaction with IMES. Conceivably, these myopic approaches are not effective in such complex and mercurial environments as IMES. In addition, there are user-based approaches that rely on ad hoc methods such as getting weights, preferences, and properties through user inputs. The usefulness of such approaches is severely limited by cognitive, informational, and functional capabilities of the user. It also distracts the user from her main objective, i.e. learning the subject at hand. Furthermore, personalization based on user inputs cannot accommodate changes in learner's interests or cognitive progress [23].

6.2 Algorithmic Approaches

An extension of deterministic approach is the assumption that some prudently devised algorithms could encompass all plans and corresponding actions. Indeed, it has been shown that plan recognition could be treated as deduction under a particular set of assumptions about the possible causes of actions [25]. Various algorithmic approaches work by determining a learner's plan from a library of possible plan schemas [9]. Such content-based algorithms may perform well at determining the general context. However, these cannot easily evaluate qualitative attributes like user's decision style, perceived usefulness, timeliness of presentation, etc. Consequently, even a good user-modeling algorithm alone does not form a truly useful system [10]. Furthermore, such algorithms require all plans to be identified to explain the learner's behavior and result in combinatorial intractability of the search task. Some probabilistic, heuristic, or soft computing approaches may reduce the search space and make plan recognition more tractable. Nevertheless, the problem in adaptive IMES is more complicated in the sense that not only we have to predict the learner's intentions but also predict learner's cognitive progress based on which system should provide higher learning opportunities.

6.3 Probabilistic Approaches

The majority of uncertainty management methodologies quantify uncertainties in form of some probabilistic measures that are propagated during reasoning [35]. Examples include the Bayesian Belief Networks, Certainty Factors, Dempster-Shafer, etc. Such approaches are based on the premise that assigning a

certain value to a plan hypothesis reflects the likelihood of its being pursued by the user [23]. Thus, it lends itself to some probability-like measure for representing information about user's individual preferences [46]. The key issue in using probabilistic approaches is accurate representation of the probabilistic dependencies in the task domain. However, student modeling using probabilistic approaches is problematic due to dynamism in student's knowledge resulting from interaction with information.

Among such traditional uncertainty modeling tools, Bayesian Networks (BN) are a popular formalism approach in user modeling and establishing sensible policies for handling uncertainty in knowledge assessment and plan recognition. BN is based on Bayes' theorem where the evidence is encoded in a directed acyclical graph with nodes corresponding to single/multi-valued variables and links corresponding to probabilistic influence relationships.

However, representing various cognitive and pedagogical aspects using BN requires assigning probabilities to such events. As such, extensive empirical experimentations are needed to establish such probabilities. Alternatively, such conditional probabilities could be estimated by domain experts or based on a more general theory about the relationships among variables of these types. However, the estimation of such probabilities by human experts is often inconsistent and biased. Past studies involving such probabilistic approaches highlight the need for highly extensive usability studies of the system involving determination of probabilities for BNs by users and experts.

Another problem with BN is that it is valid only under the simplifying assumption that the presence of evidence also affects the negation of conclusion, which is not valid in most instances. In addition, BN is not well suited for providing explanation facilities. Furthermore, the computational complexity of BN is sometimes prohibitive and representing a realistic problem solution could be quite large. In fact, it has been shown that the exact application of the Bayesian inference technique has an NP-hard nature [24]. Under dynamic conditions, the size and topology of the networks may hamper updating BN in real time. Moreover, if even a small change in the knowledge representation is required, it can affect a large number of sub-networks. Approximation techniques for applying Bayesian can be useful; however, such techniques are effective only under specified conditions.

In Certainty Factors (CF) approach, the knowledge is expressed in the form of rules and a confidence factor associated with each rule. It does not require statistical basis for supplying beliefs in events and allows simultaneous rule representation and quantification of uncertainty making it simpler and efficient compared to

BN. However, CF approach is not built on a solid theoretical foundation and results in many weaknesses such as the implicit assumption of independence among hypotheses [35].

The Dempster-Shafer (DS) theory of evidence addresses some of the weaknesses of the probabilistic approach including the representation of ignorance, the unnecessary requirement that the sum of beliefs in an event and its negation be 1 etc. [40]. DS formalism has been applied to the quantitative modeling of preferences in situations with partially or even completely missing statistical data and to compute the impact of new observations on the resulting assessment. However, it does not specify how the probabilities are to be computed or how the results are to be interpreted. Furthermore, in certain instances, obviously incorrect conclusions can be reached [2]. Moreover, the exponential nature of evidence and hypothesis spaces means application of DS is in the NP. The only way to dodge this problem is to use some heuristics to compute approximate solutions [24].

In short, such assessment of the user's action history can be helpful in establishing a numerical estimation of user's future behavior, similar to a priori probabilities over the set of all plans [15]. However, the implicit assumption of continuity in a user's attitude is clearly debatable as user behavior might change completely. Consequently, the underlying uncertainties and dynamics of the problem dictate the need for a methodology pertinent to incomplete, imprecise, inconsistent, and uncertain preferences and rules [4]. Most existing probabilistic techniques fail to deliver in uncertain environments falling in more than one of the aforementioned categories. This shortcoming is more evident and imperative when the available information is incomplete.

6.4 Machine Learning

The traditional user modeling systems have disadvantages that can be overcome with ML techniques for adaptive learning [20]. For instance, an AUI in IMES requires an ability to continuously extend the system's knowledge about the applicability and efficacy of different adaptation techniques by observing the success of such techniques in different users/contexts. The ML techniques are capable of expressing a rich variety of non-linear decision surfaces [47]. Such techniques, in general, process training/input data and attempt to make decision or classification based on this input. Furthermore, one frequently used underlying assumption in ML is the improved predictive performance by using more training data [47].

ML-based user-adaptive systems work differently from traditional Knowledge Representation (KR) based approaches. Instead of a knowledge base, observations of user behavior and history of interactions are treated

as training examples used by Learning components. The knowledge acquisition is automatic and incremental. As such, learning results are revised steadily without any special revision mechanisms [32]. Some examples of ML techniques used as alternatives for enabling adaptivity are Artificial Neural Networks, Case-Based Reasoning, Memory-based Learning, Decision Tree Induction, and Learning Automata. Multi-strategy ML approaches have also been used to engender hybrid LMs. For instance, short-term and long-term interests of user can be incorporated in a hybrid model using techniques most suitable for the specific task [10].

Although the knowledge acquisition in ML is automatic and incremental, considering the dynamic nature of user's interests, cognitive skills, and goals, the ML approaches seem inappropriate in IMES. Furthermore, the KR is implicit and formats of learning results (probabilities, decision trees, etc.) are specific to the learning algorithm. Consequently, due to lack of generalized representation formalism, ML techniques are not easily amenable to reusing of learned results for other purposes such as explanation facilities [32]. Consequently, the eventual goal of constructing a learning system that requires no intervention from the designer other than a list of potentially useful features is still elusive in realistic applications. Nevertheless, the use of ML algorithms for user modeling purposes has attracted much attention.

6.5 Fuzzy Logic

A natural way to characterize the relationship between attributes in the LM and concepts in the DM is the use of fuzzy linguistic labels [32], [33]. People often reason in terms of vague and context dependent concepts in dealing with situations where they encounter uncertainty [43]. FL techniques are used for representing and reasoning with vague concepts to mimic human style of reasoning. This reasoning may be that of the user, whose inferences or evaluations are being anticipated, or it may be that of an expert whose knowledge constitutes the basis for the system's reasoning. A user modeling system based on FL renders reasoning easy for designers and users to understand and to modify [24], [26], [33].

Furthermore, explicit information gathered from students about cognitive understanding and interests are inherently vague and subjective because they themselves might not have precise knowledge or they might not be motivated or competent to express their knowledge precisely [26]. Moreover, such information is usually incomplete. As such, the values of the attributes can be expressed in terms of linguistic labels that are handled as fuzzy numbers. Such circumstances render FL a logical choice as the membership functions of FL are in general well-suited to the representation of such input, even if the subsequent processing does not use FL techniques.

FL claims a certain degree of human-likeness because of the way in which it captures human reasoning with vague concepts [2], [3], [43]. Consequently, it seems to be a good choice for both initialization and refinement of LM as well as DM by using linguistic labels for domain concepts selected by users and/or classifying user as one belonging to various stereotypical categories. Furthermore, it sanctions tuning of the parameters of a user modeling system based on feedback from system performance, a common approach used in the field of rule-based expert systems [3], [26]. Fuzzy adaptation rules may also facilitate automated detection of conflicts and inconsistencies in the set of rules and provide robust performance in such conflicting scenarios [33].

On the flip side, it has been argued that FL was not developed for the purpose of cognitive simulation. As such, it cannot be taken for granted that an FL treatment of a given problem corresponds to the way people would deal with it [24]. Furthermore, the task of determining the appropriate representations may still require considerable empirical testing and knowledge engineering [26]. Nevertheless, the same is true with any other uncertainty management technique. Indeed, FL has been shown to be effective and robust technique in a variety of fields involving reasoning with incomplete, inconsistent, imprecise, and uncertain information [2], [4], [35]. Moreover, FL is superior to other uncertainty management techniques in terms of the computational complexity. Consequently, we believe, FL has a significant role to play in cost-effective and robust knowledge modeling under uncertain conditions as well as reasoning with such knowledge in AUI/IMES.

7. FUTURE DIRECTION IN AUI

It is not easy to predict future of AUI in IMES due to fast changing computing technology and computer users. However, we believe AUI has a very promising as well as challenging future in IMES. Here we delineate few prolific research directions in this regard.

7.1 Logical Characterization

The Functional Representation of AUI in IMES is more prevalent and hinges on various knowledge models to capture knowledge for the system. However, it renders the analysis and comparison of IMES on technical level difficult as, by and large, most existing systems are designed for special purposes [18]. Consequently, there is a need for formulating a common language [22]. Recently, a logic-based definition of IMES has been proposed allowing an abstract generalized formalization [18], [22]. Standardization of educational formalism would also facilitate the specification of many instructional models [18]. The goal is to develop a standard meta-language to be used all over the world facilitating the reuse of adaptation techniques in different contexts by reasoning over facts described in

standardized metadata formats [18], [19], [22], [30], [38]. Furthermore, reasoning can be performed in wider context over distributed data. Consequently, such representation seems to fit well for facilitating explanation facilities.

7.2 Meta-Adaptivity

As already mentioned, different adaptation techniques work most efficiently in different contexts underscoring the need for adaptation in the adaptation techniques, or meta-adaptation, during the cognitive progress of the learner. However, most empirical studies in adaptation techniques are done in simplistic and well-defined 'with or without' scenario by varying only a specific adaptation aspect and there is need for more extensive testing.

7.3 Explanation Facilities

An Explanation Facility furnishes the ability to thoroughly explore the implications of knowledge models and bases of system's adaptations. Studies have shown that a significant amount of user's sense of control is attached to the ability to readily make sense of interaction with the system. Consequently, Explanation Facilities indicating to learners, if invoked, the reasoning behind actions seems to be a nice extension to existing AUI frameworks. Furthermore, it may provide users of an AUI a sense of control by making the system 'scrutable' [17]. Such functionality may also be complemented with a visible method for controlling the system's adaptations or the user profiles. In its simplest form, explanation facilities could show rules and their usage sequence in inferences regarding certain actions [17], [35]. However, decisions regarding the effective mode for presentation of explanations would be yet another interesting research direction. Some small-scale exploratory studies have highlighted value of, as well as the difficulty in, supporting explanation/control capability because of lack of visibility or novel nature of the notion [17]. It points towards the need for making the explanation/control capabilities both visible and comprehensible to the user. Once again, simple and linguistic format of FL brings power to provide explanation facilities in a simple, compact, efficacious, and comprehensible manner [4].

7.4 Novel Scenario Modeling

When users are inactive, existing models can detect these situations but do not know the cause. Some mechanism for inferring the cause of inaction, such as lack of motivation or understanding, and adapting the interface accordingly would be worth exploring. Moreover, an interesting and more complex situation would arise when a learner switches from one access device to another during the course of learning. As such, an apparent inaction of the user could also be a result of distraction to other interesting services as well as

multi-tasking. Modeling a multi-tasking user poses even more challenging problem. Nonetheless, such a user is quite common in today's Web and Mobile service users. Furthermore, a usual classroom-teaching scenario is group/collaborative work that facilitates rich interchange of information and ideas from one another as well as from teacher [15]. It requires sensitivity to cognitive, intellectual, social, and cultural diversity. Modeling for such scenarios in adaptive IMES would certainly become a major focus in this research area.

Similarly, exploring the role of emotions in user modeling and decision-making could be a purposeful exercise [46]. There is growing evidence that emotional states may affect performance by altering perception, cognition, selection, motor actuation, etc. and may influence attention, planning, learning, memory, and decision-making. Some preliminary research work in some simplistic situations can be found in literature [15]. However, extensive work is required for incorporating emotions in IMES in realistic terms. In this regard, various theoretical frameworks may provide useful tools, such as the OCC cognitive theory of emotion and Game/Drama Theory [46],[36]. Furthermore, an interesting direction could be knowledge representation that facilitates such soft knowledge as mental and conceptual models. Once again, we believe that FL could prove a powerful modeling tool in these scenarios.

7.5 Extensive and Comparative Evaluation

The real-world deployment of user modeling and user-adapted systems with a demonstrable effectiveness is a formidable task [16], [10]. Conceivably, a review of past AUI articles reveals insufficient empirical evaluations [14]. Nevertheless, empirical studies of actual users help reinforce, contradict, and refine designs to better accommodate and satisfy users in accomplishing their tasks. However, the upward trend is quite visible and bodes well for the future of AUI and several instances of successful real world deployment of AUI systems can be cited. Nevertheless, there are problems due to the relative inability to identify evaluation needs and insufficient mappings of such needs to available resources.

Various possible metrics for evaluating an interface are: subjective evaluation of interaction quality by the user, user-friendliness, effectiveness, degree of task simplification, generalizability, scalability, accessibility, acceptability, etc. [5], [34]. However, one of the objectives in AUIs is the higher interaction quality requiring absorption of such aspects as *usability, usefulness, suitability, tailorability*, etc. that might not be measurable with currently available evaluation means and measures. Despite the difficulty in quantifying these aspects, user's rating could be useful for evaluating an AUI. Nonetheless, comprehensive evaluation instruments and techniques for guiding the

design of adaptations are largely missing. The primary reason for this is the contextual nature of adaptations that often does not facilitate objective assessments [5]. Incorporation of learner's misconceptions, beside knowledge and cognitive progress, is an important consideration for an effective AUI. Consequently, it has been argued that LM should also be validated using external tests and comparisons with actual user behavior [45]. Conceivably, all this is quite difficult to achieve.

As already mentioned, a more general description on adaptation functionality would allow easy and effective reuse of adaptation techniques in various domains or contexts. This can be achieved by making metadata about different resources explicit using standardized descriptions [18]. Consequently, empirical studies of AUIs with several dimensions of adaptivity would certainly assist in developing tools for meta-adaptation.

7.6 Generalizability, Scalability, Portability

In terms of generalization and scalability of the system, more experiments must be performed in order to determine how the system would scale up to deal with more parameters of the AUI. Knowledge representation facilitating soft knowledge, other than facts/procedures, such as mental and conceptual models is a largely unexplored direction. Furthermore, knowledge representation facilitating scaling up to larger/broader domains is an important issue. As already mentioned, this goal is significantly dependent on formalism in AUI/IMES frameworks. In addition, one of the most relevant requirements for acceptability and pervasiveness of AUI/IMES is the platform independence or portability. Such issues need to address the much-desired reusability a reality. Furthermore, in an organizational context, adaptive educational software is not only a teaching/learning resource and a carrier of instructional strategies but also a source of much touted organizational and strategic changes. Consequently, the research in AUI can be extended to incorporate organizational changes in the research framework.

7.7 Reduction in Development Time

The future research in AUI/IMES seems to be geared towards forms that are more complex and involve various multimedia embodiments. It would require articulation of abstract design patterns, elaboration and enumeration of different design alternatives, ranking of competing alternatives (based on experts' opinions and other subjective design criteria), and propagation of design knowledge into the development cycle [5]. Consequently, the need for reducing the developmental cycle time through some intelligent decision support mechanism cannot be ignored [2].

Recently, an expert system paradigm based framework has been proposed for intelligent decision support in layout design with an application to the adaptive Web

page layout design [2], [3]. We believe that such knowledge-based decision support frameworks would be valuable in fast and easy generation, evaluation, and refinement of superior alternatives. Such efforts would reduce the cognitive overload and personal bias in building a system that is rich in alternatives for various users by making use of a combination of AUI/IMES techniques/technologies, philosophies/pedagogies, etc. Furthermore, some authoring application for analyzing the set of rules by expressing just properties, without having to be concerned with the methods and techniques of inferencing, is an interesting but challenging direction. Availability of an efficient, efficant and easy-to-use off-the-shelf tool towards this would certainly make designing and utilizing AUIs more attractive. Representing knowledge, skills and contexts at a meta-level in a standardized and easily interpretable graphic language might also facilitate reuse and adaptation of models from different sources.

10. SUMMARY

The purpose of this paper is to briefly introduce the issues, prospects, and difficulties associated with research in adaptive user interfaces within the IMES context. The success of the learning process in such an educational environment depends on the intelligent context-oriented presentation of the domain knowledge by the system and its adaptivity in terms of complexity and granularity consistent to the learner's cognitive skills and progress. The richness of the human behavior and technological opportunities mean there is no final solution in such sensitive human-computer interaction area. Nevertheless, we have tried to highlight the role of various paradigms, techniques, and technologies in creating efficant AUIs that may help in providing more flexible and enhanced IMES. Furthermore, a perfunctory review of future trends and research directions has been presented. We hope that this review of complexity as well as limitations of AUI and pertinent technologies would prove helpful in creating expectations that are more realistic. However, this is just a broad overview of AUI literature and by no means complete. An integrative and comparative review of various adaptation techniques, pertinent technologies, and knowledge models with specific consideration of Web and Mobile Learning would be a worthwhile effort

REFERENCES

- [1] ACM, (2004) "ACM Code of Ethics and Professional Conduct", *ACM Computing and Public Policy: A Fair Society in Computing*, §1.4.
- [2] Ahmad, A.R., Basir, O.A., Hassanein, K., (2003) "Fuzzy Inferencing in the Web Page Layout Design", *WSMAI'03*, April'03, 33-41.
- [3] Ahmad, A.R., Basir, O.A., Hassanein, K., (2004a) "Intelligent Decision Support System for Layout Design", *Proc. of the 9th Asia-Pacific Decision Sciences Ins. Conf (APDSI'04)*, Korea, July 2004
- [4] Ahmad, A.R., Basir, O.A., Hassanein, K., (2004b) "Decision Preferences, Constraints, and Evaluation Objectives

- in Layout Design: A Review of Modeling Techniques", *Proc. of the 5th Int'l Conf. on Op. and Quan. Manag (ICOQM-V)*, Oct. 2004, Seoul, Korea, to appear.
- [5] Akoumianakis, D., Savidis, A., et al., (2000) "Encapsulating intelligent interactive behaviour in unified user interface artefacts", *Interacting with Computer*, Vol. 12, 383-408.
- [6] Alpert, S.R., Karat, J., Karat, et al. (2003) "User Attitudes Regarding a User-Adaptive eCommerce Web Site", *UMUAI*, Vol. 13, 373-96.
- [7] Benyon, D.R. (1996) "Domain Models in User Interface Design", *Critical Issues in User Interface Systems Engineering*, Springer-Verlag.
- [8] Bianchi-Berthouze, N., L. Berthouze, and T. Kato, (1999) "Understanding Subjectivity: An Interactionist View", *UM'99*, Springer, NY, 3-12.
- [9] Bikel, D., Schwartz, R., Weischedel, R., (1999) "An algorithm that learns what is in a name", *Machine Learning*, 34(1), 211-231.
- [10] Billsus, D., Brunk, C.A., Evans, C., Gladish, B., Pazzani, M. (2002) "Adaptive interfaces for ubiquitous Web access", *Comm. of the ACM*, 45(5), 34-38
- [11] Bontcheva, K., (2002) "Adaptivity, Adaptability, and Reading Behaviour: Some Results from the Evaluation of a Dynamic Hypertext System, *AH'2002*.
- [12] Brusilovsky, P., (2001) "Adaptive hypermedia", *UMUAI*, 11(1), 87-110.
- [13] Brusilovsky, P. (2003a) "Adaptive navigation support in educational hypermedia: the role of student knowledge level & case for meta-adaptation", *Brit. J. of Edu. Tech*, 34(4), 487-97.
- [14] Chin, D.N., Crosby, M.E., (2002) "Introduction to the Special Issue on Empirical Evaluation of User Models and User Modeling Systems", *UMUAI*, Vol. 12, 105-109.
- [15] Conati C., Zhao X. (2004) "Building and Evaluating an Intelligent Pedagogical Agent to Improve the Effectiveness of an Educational Game", *IUI'04*, Jan. 13-16, Madeira, Portugal.
- [16] Corbett, A., McLaughlin, M., Scarpinato, K., (2000) "Modeling Student Knowledge: Cognitive Tutors in High School & College", *UMUAI*, Vol. 10, 81-108.
- [17] Czarkowski, M., Judy, K. (2003) "How to give user a sense of control over personalization", *AH'03*, 121-132.
- [18] Dolog, P., Henze, N. (2003) "Personalization Services for Adaptive Educational Hypermedia", *ABIS 03 Workshop*, Oct. 6-8, Karlsruhe, Germany.
- [19] Educational Modeling Language (EML) Group: <http://eml.ou.nl/>
- [20] Goecks, J., Shavlik, J., (2000) "Learning Users Interests by Unobtrusively Observing Their Normal Behavior", *IUI'00*, New Orleans, LA.
- [21] He S., Kinshuk, Hong H., Patel A. (2002), "Granular Approach to Adaptivity in Problem-based Learning Environment", *ICALT'02*, Palmerston North, NZ, 3-7.
- [22] Henze, N., Nejdli, W., (2003) "Logically Characterizing Adaptive Educational Hypermedia Systems", *AH'03*, Budapest, Ungarn.
- [23] Jameson, A. (2001) "Modelling both the context & user", *Personal Technologies*, 5(1), 1-4.
- [24] Jameson, A., (1996) "Numerical uncertainty manag. in user/student modeling: An overview of systems & issues", *UMUAI*, Vol.5, 193-251.
- [25] Kautz, H.A., (1987) "A formal theory of plan recognition", *PhD thesis*, U. of Rochester, NY.
- [26] Kavčič, A., Jiménez, R.P., et al. (2003) "Student Modelling Based on Fuzzy Inference Mechanisms", *Proc. of EUROCON'03: Computer as a Tool*, Ljubljana, Slovenia, Sep. 2003
- [27] Kinshuk, and Patel, A. (2001) "Implementation issues in web-based training", *Web-Based Training*, Educational Technology Publ., NJ.
- [28] Kinshuk & Lin T., (2003) "User Exploration Based Adaptation in Adaptive Learning Systems", *Int'l J. of Information Systems in Education*, 1(1), 22-31.
- [29] Kobsa, A., (2003) "Privacy through pseudonymity in user-adaptive systems", *ACM Trans. on Internet Technology*, 3(2), 149-183.
- [30] Learning Object Metadata (LOM), <http://ltsc.ieee.org/doc/wg12/LOM-WD3.html>
- [31] Maes, P., (1994) "Agents that reduce workload", *Comm. of the ACM*, 37(7), 31-40.
- [32] Maleki-Dizaji, S., Othman, Z. A., et al. (2003) "Evolutionary Reinforcement of User Models in an Adaptive Search Engine", *WI'03*, 706-9.
- [33] Martinovska, C., (2002) "A Fuzzy-Based Approach to User Model Refinement in Adaptive Hypermedia System", *AH'2002*, 411-414.
- [34] McTear, M.F., (2000) "Intelligent Interface Technology: From Theory to Reality", *Interacting with Computers*, Vol.12, 323-336.
- [35] Negnevitsky, M., (2002) *Artificial Intelligence: A Guide to Intelligent Systems*, Pearson.
- [36] Obeidi, A., K.W. Hipel, D.M. Kilgour, (2003) "Emotion: The Missing Ingredient in Conflict Analysis", *Int'l Conf. on Systems, Man and Cybernetics*, Vol. 4, 3322 - 3329.
- [37] Sampson D., Karagiannidis C., Kinshuk (2002), "Personalised Learning: Educational, Technological and Standardisation Perspective", *Interactive Edu. Multimedia*, Vol. 4, 24-39.
- [38] SCORM: The Sharable Content Object Reference Model, <http://www.adlnet.org/scorm/>.
- [39] Serafini, L., Bouquet, P. (2004) "formal theories of context in AI", *AI*, Vol. 155, 41-67.
- [40] Shafer, G., (1976) *A Mathematical Theory of Evidence*. Princeton University Press, NJ.
- [41] Shneiderman, B. (1997) "Direct manipulation for comprehensible, predictable and controllable user interfaces", *IUI'97*, Orlando, FL.
- [42] Shneiderman, B. Maes, P. (1997) "Direct manipulation vs. interface agents", *Interactions*, 4(6), 42-61.
- [43] Turban, E., Aronson, J.E., (2001) *Decision Support & Intelligent Systems*, PrenticeHall, NJ.
- [44] Vomlel, J. (2004) "Bayesian Networks in Educational Testing", *IJUFKS*, 12(1), 83-100.
- [45] Weibelzahl, S., Weber, G. (2003) "Evaluating the Inference Mechanism of Adaptive Learning Systems", *LNCS*, Vol. 2702, 154-162.
- [46] Zhou X. and Conati C. (2003) "Inferring User Goals from Personality and Behavior in a Causal Model of User Affect", *IUI'03*, Miami, FL, 211-218.
- [47] Zukerman, I., Albrecht, D., (2001) "Predictive Statistical Models for User Modeling", *UMUAI*, 11(1), 5-18..

Computer-Mediated Communication and Innovation: Do Communication Media Properties Influence Innovative Thinking Processes?

Leif Jarle Gressgård, Gunnar E. Christensen

NHH, Department of Strategy and Management, N-5005 Bergen, Norway.
leif.gressgard@nhh.no, gunnar.christensen@nhh.no

ABSTRACT

This paper concerns potential impacts that characteristics of communication media in computer-mediated problem solving in groups have on group-members' innovative thinking. Different processes involved in effective idea generation and problem solving are presented, and we discuss how three characteristics of communication media relevant for group collaboration aiming at producing innovative ideas might influence the group process. A research model is presented, and hypotheses regarding the effects of affordances on innovative thinking are put forth. We finally describe a possible methodological approach that might be applied in order to test the proposed hypotheses.

Key words: CMC, Computer-mediated communication, Innovative thinking, Problem solving, Idea generation.

1. INTRODUCTION

Superior performance in product development and innovation is believed to be one of the main sources of competitive advantage in the modern market place, and information technology used to facilitate innovative and creative processes is used in a variety of organizational settings. In this business environment, where innovation processes are facilitated by computer-mediated communication, it is important to obtain knowledge of *whether (and in case how) qualities or properties of electronic communication media influence individuals' innovative thinking in group-based problem solving*. This research question is addressed in this paper.

Many studies in the computer-mediated communication (CMC) literature focus on differences between various "communication modes" (e.g. face-to-face and dispersed CMC), without digging deeper into the underlying qualities of the modes. However, in order to generate new knowledge on desirable use of ICT in groups, it is important to investigate how the distinctive characters of different CMC settings impact on relevant dependent variables. We will therefore put emphasis on variables that all mediated and non-mediated communication processes can be described and evaluated by in this paper. There are many theories that have been developed in this research stream, and accordingly there are many theoretical concepts describing underlying features of communication modes. Theories like Media Richness Theory [12], Media Synchronicity Theory [14] and Burgoon et al.'s [6,7] interactivity model all make efforts in describing the constituent parts of mediated communication. Based on these contributions, we discuss media properties that may be of particular importance for innovation processes in group-based problem solving. In studies occupied with the relationship between use of ICT in group interaction and innovation, the dependent (innovation) variables investigated are most

often the products or artifacts resulting from the group processes. Variables often found in this research domain are for example the number and quality of the ideas generated [e.g. 10]. Accordingly, little emphasis has been put on how communication media characteristics influence the *processes* which lead to better or more desirable scores on the before mentioned outcome variables. This research applies such a process perspective, and focus on innovative thinking processes for each individual in a group-based problem solving situation. The paper is organized as follows: We first present and discuss the theoretical constructs that are relevant for our study, then we present the research model and our hypotheses. In the final part we present a potential methodological approach for testing the hypothesized relationships.

2. THEORY

2.1 Innovative Thinking in Idea Generation

Innovation can be understood as a process of creating or modifying an idea and developing it to produce products, services, processes, structures, or policies that are new to the organization [24]. The phases in the innovation process are conceived to encompass the generation, development, and implementation of new ideas and behaviors [5]. This conceptualization of innovation is highly related to *organizational creativity*, which can be defined as "the creation of a valuable, useful new product, service, idea, procedure, or process by individuals working together in a complex social system" [23 p. 293]. Both innovation and organizational creativity are initiated with idea generation, and successful outcome of this phase is related to certain thinking processes. The concept of *thinking* is often construed as "an umbrella term for a range of *processes* associated with "high-level" *cognition*, such as reasoning, categorization, and judgment and decision making" [17 p. 266, our italics]. In a problem solving situation, the first process that

needs to be initiated is a search for alternatives. That means that the decision makers or problem solvers have to collect information that is relevant to the problem, and which may contribute to, or are necessary for, a successful change. After the search for alternatives, the decision makers need to select one specific solution that they find most appropriate for the task they face. There exists little controversy regarding the importance of these two “opposite” phases to be present in idea generation activities; in other words, successful completion of idea generation entails a combination of *divergent* and *convergent* thinking.

2.1.1 The Roles of Divergent and Convergent Thinking in Idea Generation

In a problem solving context, exercising divergent thinking involves starting with a specific problem and generating various options and perspectives on the problem. Convergent thinking follows the divergent process, and acts to narrow down the options available to obtain a number of “satisfying” solutions to the problem (figure 1).

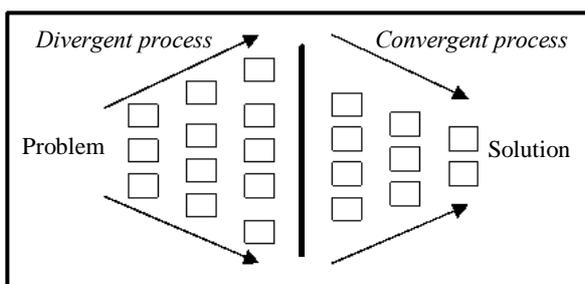


Figure 1: Divergent and convergent thinking

In this context, divergent thinking refers to going off in new directions rather than thinking solely on one solution, and deriving a variety of ideas from given information. Following divergent thinking, the ideas and information will be organized using convergent thinking; i.e., putting the various ideas back together in some organized, structured way. Some authors advocate the deliberate distinction between divergent and convergent thinking processes. Belonging to this school of thought, Basadur et al. [4] identified a sequenced two-step thinking process called “ideation-evaluation”. They defined “ideation” as the production of ideas without evaluation, and “evaluation” as the application of judgment to the ideas produced. One of the reasons for making a clear distinction between these processes is to avoid that people hold back ideas they think are stupid or silly, and by this using divergent and convergent thinking interchangeably. To create significant improvements or entirely new products, services or processes, those ideas that seem absolutely preposterous or unachievable at first are needed. Empirical research has supported both the general separation of idea production from idea

selection and the more specific ideation-evaluation process [1,3,4].

Basadur and Finkbeiner [3] view ideation as having both *cognitive* and *attitudinal* elements. Effective ideation may thus “require specific attitudes favoring this kind of thinking, perhaps to help participants truly “let loose” and use more fully their unencumbered imaginations”(p. 38). Such attitudes may for example impact on decisions of whether ideas that are produced should be put forth, and are therefore important aspects of the idea generation process. This is consistent with several “general” theories concerning the linkage between attitudes and behavior, like Kraut’s [18] training model suggesting a causal chain whereby attitude change leads to performance change, and Fishbein and Azjen’s [15] theory of reasoned action proposing that behavior can be predicted by individual attitudes and social norms.

2.1.2 Attitudes Toward Idea Generation

According to Basadur et al. [4] and Basadur and Finkbeiner [3], persons have attitudes toward both ideation and evaluation, and they have identified several *attitudinal constructs* related to idea generation. Two of these constructs are “preference for ideation” and “tendency for premature critical evaluation of ideas”. The denotation of the first construct concerns mind-sets such as “being less likely to jump to conclusions as to what is the real problem”, and “more open-minded to new ideas and approaches [2 p. 22], while the mind-sets associated with the latter construct are more or less opposite¹. Basadur and Finkbeiner developed internally valid and reliable measures of these constructs, which we will adopt in our study.

2.2 Communication Media: Capabilities

Many theories have been developed to categorize media qualities and explain media effects on communication outcomes [e.g. 16]. One of the most widely used media categorization theories is *Media Richness Theory* [12], which suggests that communication media can be ranked on a richness-leanness continuum, and asserts that task performance will be improved when tasks’ information needs are matched to a medium’s richness or it’s capabilities to facilitate shared meaning. Empirical tests have not been very supportive, however, particularly for new communication media [e.g. 8, 13]. Recognizing the weaknesses of Media Richness Theory, Dennis and Valacich [13] introduced *Media Synchronicity Theory*, which suggest that media have five capabilities that are important in understanding the effects of media use on the ability to communicate and process information; 1) immediacy of feedback, 2) symbol variety, 3)

¹ As “tendency for premature critical evaluation of ideas” is more or less antagonistic to “preference for ideation”, we use the term “preference for evaluation” for this particular attitude.

parallelism, 4) rehearsability, and 5) reprocessability. Both media richness theory and media synchronicity theory categorize media attributes in terms of what kind of properties or capabilities the different media possess. Following this kind of reasoning, Burgoon et al. [6,7] suggest two ways of characterizing the *interactivity* concept often used in discussions of CMC. First, it can be considered in terms of the structural properties of the medium, or second, by qualitative experiences as perceived by the user. Regarding the first categorization type, Burgoon et al. integrate the various analysis of media affordances into an extensive set of properties, consisting of: 1) *participation* (extent of engagement in the interaction), 2) *mediation* (whether the interaction is mediated or not), 3) *contingency* (extent to which contributions are dependent on prior postings), 4) *media and information richness* (whether the format utilizes one or more modalities, and the extent to which it supports symbol variety), 5) *geographical propinquity* (physically co-located or distributed participants), 6) *synchronicity* (same-time or asynchronous interaction), 7) *identification* (fully identified, partially identified or anonymous participants), 8) *parallelism* (whether the format permits concurrent communication and multiple addressees, or only permits serial messages), and 9) *anthropomorphism* (extent to which the system interface simulates or incorporates humanlike characteristics). Communication media can thus be described and evaluated in terms of their abilities to enable specific aspects or properties of the communication process. In other words, all communication processes can be characterized by certain *affordances*, and communication media differ in their capabilities to support and enable these affordances. We will place emphasis on three affordances that may influence attitudes toward idea generation: “synchronicity”, “parallelism”, and “identification”.

2.2.1 Synchronicity

Synchronicity refers to whether the interaction is same-time or not. A high level of synchronicity enables the interacting parties to give immediate feedback, which refers to the extent to which a medium enables users to give rapid responses to the communications they receive [12]. Opportunity for rapid bidirectional communication has been shown to affect communication outcomes by increasing interaction between the parties, allowing rapid assessment and modification of the message [25]. We thus define synchronicity as “the extent to which the participants engaged in a problem solving situation can give immediate feedback to the postings of other group members, and receive immediate feedback on their own postings from other group members”. We are concerned with the level of synchronicity *as perceived by the participants* in a group problem solving situation (as people may have different perceptions of the same

technology). The highest level of synchronicity is most likely to occur when interaction takes place in real time, and the level will decrease along with increasing time lag between messages from one participant and responses to these messages from another participant. It is important to emphasize that a high level of synchronicity does not assure immediate feedback; it is rather related to the *opportunities* that the participants have for giving rapid responses to the messages of others. Low synchronicity thus implies that the participants involved in interaction do not have the opportunity to give and receive immediate feedback.

2.2.2 Parallelism

Many electronic media can be structured to enable multiple interacting parties in one session, and parallelism refers to the number of simultaneous conversations that can *effectively* take place. High parallelism can increase the amount of information that can be simultaneously transmitted and received, but it can also decrease the effectiveness of information processing as it may lead to information overload. We define parallelism as “the participants’ opportunities to be engaged in simultaneous dialogues in a group problem solving situation”. High parallelism thus gives the participants opportunities to be engaged in multiple dialogues at the same time, and therefore they do not have to take turns in utterance of contributions. Low parallelism, on the other hand, means that all participants have to be engaged in a single dialogue, implying that only one member of the group can utter his/her ideas and comments at the same time. We are also for this affordance concerned with the levels as perceived by the participants in a group problem solving situation.

2.2.3 Identification

Individuals involved with idea generation may be influenced by the presence of others. Persons are concerned about how others perceive their ideas [10, 19,20], and depending on this evaluation apprehension, embedding idea generation in a group setting could either enhance or reduce an individual’s contribution of ideas. On the positive side, the group might provide encouragement, stimulation or reward for creative contributions, and on the negative side, contributors might anticipate embarrassment, hostile evaluation, conformity pressures or other punishments for proposing unusual ideas [9]. Several authors have identified the anonymity option as a particular virtue of Group Decision Support Systems in that anonymity encourages full participation of group members that otherwise would have been socially inhibited from expressing unpopular, novel or heretical opinions [21]. We define identification as “the extent to which the contributor of ideas, comments, etc. in a group problem solving situation is known by the other participants”. Similar to the other affordances, we are also in this

case interested in the level of identification as perceived by the participants.

Based on the theoretical review, we propose a research model that consists of “synchronicity”, “parallelism”, and “identification” as the independent variables, and “preference for ideation” and “preference for evaluation” as the dependent variables (figure 2).

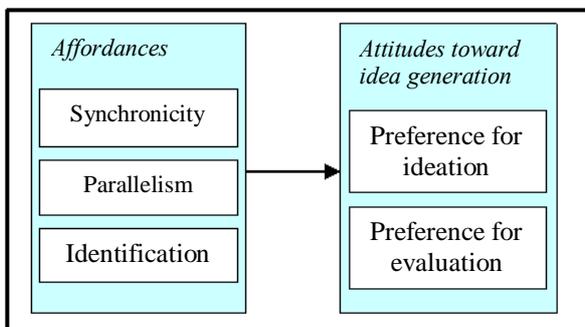


Figure 2: Research model

3. HYPOTHESES

3.1 Effects of affordances on attitudes toward idea generation

Information and communication technologies differ in their abilities to support affordances of the communication process, and synchronicity, parallelism, and identification may contribute differently to innovative thinking. We therefore expect that affordances will have different effects on attitudes toward idea generation.

3.1.1 Synchronicity

Feedback is a powerful tool for innovation processes, and it is important to emphasize the role of immediacy of feedback (timing) as a particular collaboration characteristic in idea generation. That is, timing of feedback is important in order to avoid a situation where evaluation stifles innovation before it has a chance to develop. While there are many different methodologies about how to most effectively engage in idea generating activities, one element that they have in common is that the free flow of initial ideas must occur without the interruptions of criticisms or evaluations. Low synchronicity enables both rehearsability and reprocessability, which may be regarded as “evaluative activities”. High synchronicity on the other hand, does not render possible a critical examination of neither the ideas that the sender is to put forth (rehearsability), nor the ideas or messages that an individual has received from other participants (reprocessability) before composing a response. Thus, interaction in same-time, although enabling interruption and immediate critical feedback, may have positive impacts on ideation. On this basis, the following hypothesis is put forth:

H1: High degree of synchronicity correlates positively with “preference for ideation”, and low degree of synchronicity correlates positively with “preference for evaluation”.

3.1.2 Parallelism

According to Van de Ven et al. [22], an increase in the number of initiatives undertaken by a large number of interacting people enhances the probability of stimulating innovation. With reference to the distinction between ideation and evaluation, one aspect that support the assertion put forward by Van de Ven et al., is that high degree of parallelism makes it difficult to reprocess ideas that have been put forth, while at the same time being attentive to the ongoing discussion. Hence, there is no room for a comprehensive critical evaluation of all ideas that are posed. Another important aspect of parallelism regarding idea generation, which may be a consequence of the absence of opportunities for critical evaluation as discussed above, is that it can create an environment of interaction and discussion that facilitates hitchhiking. However, there will not be an exponential increase in hitchhiking effects with increasing parallelism. The reason for this is that individuals cognitive capacities are limited, thus there may be a problem of information overload if the number of participants exceeds a certain limit. Conversely, low (or absence of) parallelism will probably result in more (and critical) rehearsability of the ideas that the participants are to suggest. In these cases, the participants are aware of that they have the other participants’ full attention when contributing their ideas, and are thus likely to be more self-critical when communicating. The following hypothesis may be stated:

H2: High degree of parallelism correlates positively with “preference for ideation” and low degree of parallelism correlates positively with “preference for evaluation”.

3.1.3 Identification

Participants in joint idea generating tasks are influenced by each other. This is particularly relevant when it comes to the participants’ decisions of whether or not to contribute and express their ideas. Most individuals are concerned about how others perceive and think of them, and can therefore be reluctant to express unorthodox or non-conforming ideas. Working in a group where the identities of the participants are known might inhibit a contributor who anticipates embarrassment, hostile evaluation, conformity pressures, or other punishments for proposing unusual ideas. In contrast, anonymity may lead to a reduction of these mechanisms. That is, it is reasonable to believe that the fear of getting negative comments will be lower if the interacting parties do not know the identity of each other. In other words, the barriers related to

idea expression that people experience, will be lower if the proposals can be done anonymously. Thus:

H3: Unknown identities (anonymity) correlate positively with "preference for ideation" and known identities correlate positively with "preference for evaluation".

4. METHOD

The focus of the study is on whether/how communication media influence individuals' innovative thinking processes within group settings. Thus there are three "key components" that the methodology must be based on in order to answer the research question: 1) The setting must involve group collaboration, 2) the participants must be involved in a joint problem solving situation, where the focus is placed on the participants' thinking processes, and 3) they must interact by use of electronic media (collaborative software / tools).

4.1 Research Design

The design of the study must ensure that the measurement of the variables is adequate, and that the variance of affordances is sufficient for hypotheses testing. Further, when hypotheses suggesting a correlation between variables are posited, the research design must control for alternative explanations. In such situations, internal validity must be given priority [11], and an experimental design is therefore preferable. Subjects participating in the experiments will be randomly distributed into groups that will be given tasks which necessitate CMC between the members in order to be accomplished. Within the experimental groups, the affordances will be manipulated, and attitudes toward idea generation will be measured on an individual level. Each group will consist of 3-5 participants, and 2 groups will be needed for determining the effects of each affordance. As the objective of this study is to investigate the effects of specific affordances on attitudes toward idea generation, and not on revealing differences in effects of non-mediated versus electronically mediated interaction on these attitudes, we find it unnecessary to include FtF interaction as an experimental condition in the study. The design we will apply is illustrated in figure 3. When focusing on synchronicity, we compare the aggregate synchronicity-score in groups 2 and 3

	Group 1	Group 2	Group 3
Synchronicity	Low	High	
Parallelism	High	Low	High
Identification	Identified		Anonymous

Figure 3: Experimental design with the synchronicity-score in group 1. The same process is done for the other affordances as well

(groups 1 and 3 versus group 2 for parallelism, and groups 1 and 2 versus group 3 for identification).

4.2 Tool kit and manipulation of affordances

The experimental conditions represent our manipulations of the independent variables. We will use specific Internet-based software for group collaboration (Groove 2.5), which has a set of basic functions that form the basis for the interaction between group members. For all experimental groups, the technology will be based on the same "framework", with the exception of the specific configuration that is needed in order to provide for variation of affordances. In the next sections we describe how the variables will be manipulated.

Based on the definition of synchronicity proposed in section 2.2.1, we operationalize this variable as *the time that elapses from the messages of one participant are posted to these messages are received by another participant*. For groups 2 and 3, we will ensure (perceived) high synchronicity by letting the participants work online. By using the shared space in an online condition, all participants in a group will have the same user interface, meaning that the contributions of one participant will pop up immediately on the other participants' computer screens. In contrast, the participants in group 1 will work offline. During the one problem solving task, the participants will log into the shared group space every three minutes (one problem solving task lasts for 30 minutes). These online periods will be as short as possible, only enabling the work space to be updated.

The operational definition of parallelism we will use is as follows: *The degree to which it is possible for the participants to post their ideas and comments at the same time without interrupting others*. In the shared work spaces of groups 1 and 3, the level of parallelism should be high. Implementing this characteristic means that all participants in the groups will be able to post their contributions simultaneously without interrupting others. Conversely, the contributions of the participants in group 2 must be posted one at a time in order to avoid interruption of others. It is worth notice, however, that the collaborative software does allow for simultaneous postings in the same window, but this type of interaction will not be efficient (this will be the same as when people speak all at once in an FtF setting).

Based on the definition of identification proposed in section 2.2.3, this variable is operationalized as *whether or not the identities of the participants in a group problem solving situation are displayed along with (and thus linked to) the contributions of the participants*. In order to ensure *perceived identification*, the names of the contributors will be displayed together with their postings, meaning that the contributions can

be identified. The identities of the participants (both online and offline) in the shared space are also shown in a “list of participants”. Contrary, in group 3, the identity of the participants will not be displayed together with their contributions and feedback, and the participants will not know the identities of the other group members.

REFERENCES

- [1] Basadur, M., S. Taggar, P. Pringle, “Improving the Measurement of Divergent Thinking Attitudes in Organization”, *The Journal of Creative Behavior*, Vol. 33, No. 2, pp75 –111, 1999.
- [2] Basadur, M., P.A. Hausdorf, “Measuring Divergent Thinking Attitudes Related to Creative Problem Solving and Innovation Management”, *Creativity Research Journal*, Vol.9, No.1, pp21–32, 1996.
- [3] Basadur, M., C.T. Finkbeiner, “Measuring Preference for Ideation in Creative Problem-Solving Training”, *The Journal of Applied Behavioral Science*, Vol.21, No.1, pp37–49, 1985.
- [4] Basadur, M.S., G.B. Graen, S.G. Green, “Training in Creative Problem Solving: Effects on Ideation and Problem Finding and Solving in an Industrial Research Organization”, *Organizational Behavior and Human Performance*, Vol. 30, 41 –70, 1982.
- [5] Blazevic, V., A. Lievens, “Learning During the New Financial Service Innovation Process: Antecedents and Performance Effects”, *Journal of Business Research*, Vol. 55, pp1 - 18, 2002.
- [6] Burgoon, J.K., J.A. Bonito, B. Bengtsson, C. Cederberg, M. Lundberg, L. Allspach, “Interactivity in Human-Computer Interaction: A Study of Credibility, Understanding, and Influence”, *Computers in Human Behavior*, Vol. 16, No. 6, pp553 - 574, 2000.
- [7] Burgoon, J.K., J.A. Bonito, B. Bengtsson, A. Ramirez, N.E. Dunbar, N. Miczo, “Testing the Interactivity Model: Communication Processes, Partner Assessments, and the Quality of Collaborative Work”, *Journal of Management Information Systems*, Vol.16, No.3, pp33-56, 2000.
- [8] Burke, K., L. Chidambaram, “How Much Bandwidth is Enough? A Longitudinal Examination of Media Characteristics and Group Outcomes”, *MIS Quarterly*, Vol. 23, No. 4, pp557 –580, 1999.
- [9] Collaros, P.A, L.R. Anderson, “Effects of Perceived Expertness upon Creativity of Members of Brainstorming Groups”, *Journal of Applied Psychology*, Vol. 53, pp159 –163, 1969.
- [10] Connolly, T., L.M. Jessup, J.S. Valacich, “Effects of Anonymity and Evaluative Tone on Idea Generation in Computer-Mediated Groups”, *Management Science*, Vol.36, No.6, pp689–703, 1990.
- [11] Cook, T.D., D.T. Campbell, *Quasi-Experimentation. Design & Analysis Issues for Field Settings*, Houghton Mifflin Company, Boston, 1979.
- [12] Daft, R.L., R.H. Lengel, “Organizational Information Requirements, Media Richness and Structural Design”, *Management Science*, Vol. 32, No. 5, pp554 - 571, 1986.
- [13] Dennis, A.R., S.T. Kinney, “Testing Media Richness Theory in the New Media: Cues, Feedback, and Task Equivocality”, *Information Systems Research*, Vol.9, No.3, pp256–274, 1998.
- [14] Dennis, A.R., J.S. Valacich, “Rethinking Media Richness: Towards a Theory of Media Synchronicity”, *Proceeding of the 32nd Hawaii International Conference on Information Systems*, Maui, 1999.
- [15] Fishbein, M, I. Ajzen, *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*, Addison-Wesley, Reading, MA, 1975.
- [16] Fulk, J., B. Boyd, “Emerging Theories of Communication in Organizations”, *Journal of Management*, Vol. 17, No. 2, pp407 –446, 1991.
- [17] Holyoak, K.J., B.A. Spellman, “Thinking”, *Annual Review of Psychology*, Vol. 44, No. pp265 – 315, 1993.
- [18] Kraut, A.I. “Developing Managerial Skills via Modeling Techniques: Some Positive Research Findings - A Symposium”, *Personnel Psychology*, Vol. 29, pp325 –328, 1976.
- [19] Nunamaker Jr, J.F, A.R. Dennis, J.S. Valacich, D.R. Vogel, “Information Technology for Negotiating Groups: Generating Options for Mutual Gain”, *Management Science*, Vol. 37, No. 10, pp1325 – 1346, 1991.
- [20] Nunamaker Jr., J.F., A.R. Dennis, J.S. Valacich, D.R. Vogel, J.F. George, “Electronic Meeting Systems to Support Group Work”, *Communications of the ACM*, Vol. 34, No. 7, pp40 –61, 1991.
- [21] Nunamaker, J.F., L.M. Applegate, B.R. Konsynski, “Computer-Aided Deliberation: Model Management and Group Decision Support”, *Operations Research*, Vol. 36, No. 6, pp826 –848, 1988.
- [22] Van de Ven, A.H., D.E. Polley, R. Garud, S. Venkataraman, *The Innovation Journey*, Oxford University Press, New York, 1999.
- [23] Woodman, R.W., J.E. Sawyer, R.W. Griffin, “Toward a Theory of Organizational Creativity”, *Academy of Management Review*, Vol. 18, No. 2, pp293 –321, 1993.
- [24] Zhuang, L. “Bridging the Gap Between Technology and Business Strategy: A Pilot Study on the Innovation Process”, *Management Decision*, Vol. 33, No. 8, pp13 –21, 1995.
- [25] Zmud, R.W., M.R. Lind, F.W. Young, “An Attribute Space for Organizational Communication Channels”, *Information Systems Research*, Vol.1, No.4, pp440– 457, 1990

Corporate Mission and Vision and Customer Knowledge Management for Increasing Competitive Advantages

Hwanyann Su ¹, Yichen Lin ², Chiu-hua Lai ³

¹ Department of Asia-Pacific Industrial and Business Management, National University of Kaohsiung, Kaohsiung, 811, Taiwan, China, h_su@nuk.edu.tw

² Graduate Institute of Technology Management, National University of Tainan, Tainan, 700, Taiwan, China

³ Department of Management and Information Technology, Southern Taiwan University of Technology, Tainan, Taiwan, China

ABSTRACT

Customer knowledge is highly valued in today's business environment where supply is more than demand and firms are trying their best to differentiate themselves from their competitors. Building customer knowledge base through knowledge management provides firms with strategic weapon for customer value creation, average level of customer knowledge enhancement, cost and time reduction, service quality improvement and quicker customer knowledge creation. The main objective of this study is to propose a conceptual framework for exploring the relatedness of the three dimensions: resource provision, knowledge management process and customer knowledge performance. The proposed conceptual framework emphasizes the link between knowledge management with corporate mission and vision and that they set out the direction of corporate knowledge management. This study takes a kitchenware manufacturer and seller in Taiwan as an example to perform a field study to collect the concerned data for exploring the expressed relatedness among the three dimensions of the proposed framework as reported by the respondents in the case company. From the responses of the surveyed case company, it is also demonstrated how customer knowledge can be enhanced through resource provision and knowledge management and that people who surely understand their corporate mission and vision can outperform in knowledge management process and in customer knowledge performance compared with people who do not in this particular setting.

Keywords: customer knowledge, knowledge management, competitive advantages, vision, mission.

1. INTRODUCTION

The focus of global economics has shifted from labor-intensive or capital-intensive industries to technology-intensive and knowledge-intensive industries. Davenport and Prusak (1998) point out that knowledge has become the major asset of modern businesses and the key to retain their competitiveness. Nonaka (1991) also stresses that in the future in which the only thing that can be certain is high uncertainty, knowledge is a certain and sustainable source of competitive advantages. Platt, the CEO of HP, also indicates that if HP can know exactly what knowledge it has owned, its profit can be increased to three times of its current level (Davenport and Prusak, 1998). Therefore, corporate knowledge management has become a very important issue for modern businesses to retain and strengthen competitiveness in the era of knowledge economics.

Among all the knowledge that needs to be managed, customer knowledge is highly valued in today's business environment where supply is more than demand and firms are trying their best to differentiate themselves from their competitors. Building customer knowledge base through knowledge management provides firms with strategic weapon for customer value creation, average level of customer knowledge enhancement, cost and time reduction, service quality improvement and quicker customer knowledge creation.

Under severe global competition, ways of retaining and strengthening competitiveness in the era of knowledge economics by means of building customer knowledge base through corporate knowledge management need further investigation. A conceptual framework for managing customer knowledge and a case study that intend to investigate how this can be done is thus provided. This paper will proceed as follows: section 2 will discuss the background of the research; section 3 will present the research design and method; section 4 will provide the research findings, followed by section 5 to further discuss them; section 6 will summarize the paper.

2. BACKGROUND TO THE RESEARCH

Corporate vision refers to the desired future state of a business (Parikh and Neubauer, 1993; Schoemaker, 1992). Corporate mission refers to its purpose, why it exists, core competencies and who benefits from its operation (Brien and Meadows, 2000). Clear mission and vision enable people in a business to have clear objectives, sense of direction, priority, holistic consideration, co-ordinated actions and the basis of performance measures. Corporate mission and vision also choose for a business its current problems and those in its future, for it recognizes them when its current state significantly deviates from its mission and vision.

Corporate mission and vision should guide and regulate corporate knowledge management at a corporate level. Arthur Andersen Business Consulting and APQC (1996) stress that corporate vision, value and strategy set out a direction for a business regarding its knowledge management. Therefore, without the linkage of corporate knowledge management with corporate mission and vision, corporate knowledge tends to be managed for its own sake.

2.1. The important resource provision for corporate knowledge management

From the survey of literature, this paper proposes two major resource inputs for successful corporate knowledge management: "incentives" and "organization and technology". These two resource inputs are detailed below.

1. Incentives: Knowledge is the source of power. A business cannot expect its employees to contribute their knowledge and share it with others without asking for rewards. A business should try to provide better treatment for those who contribute and share their knowledge than for those who do not. Stenmark (2001) also points out that knowledge sharing needs incentives and needs to make clear what rewards can be obtained.
2. Organization and technology: With delegated organization, delegated people will be responsible for the integration and co-ordination of inter-departmental knowledge management activities. Some well-known knowledge management case studies such as Arthur Andersen Business Consulting (Arthur Andersen Business Consulting and APQC, 1996) and Buckman Laboratories (Pan and Scarbrough, 1998) also set up delegated knowledge management departments. Regarding technology, related researches in the literature have clearly indicated that information technology alone cannot encourage the sharing of knowledge (Davenport, 1994, 1997; Cross and Baird, 2000). However, information technology and systems still play a very important instrumental role in the era of information and knowledge economics (Arthur Andersen Business Consulting and APQC, 1996; Davenport and Prusak, 1998).

2.2. Knowledge management process

Beckman (1997) points out that there are eight steps of a knowledge management process: Knowledge Define, Knowledge Capture, Knowledge Select, Knowledge Store, Knowledge Share, Knowledge Apply, Knowledge Create and Knowledge Sell. The "Define" function of knowledge management process is to identify the core competence of a business and the related knowledge categories, demands and requirements needed by the identified core competence. The "Capture" function is to capture from the internal and external source of knowledge the needed knowledge and to document it. The "Select" function is to select among the captured

knowledge that which is appropriate. The "Store" function is to properly codify and store the selected knowledge. The "Share" function is to disseminate and transfer knowledge. The "Apply" function is to use, integrate and reuse knowledge to carry out tasks and solve problems. The "Create" function is to create new knowledge and to uncover the existing knowledge. The "Sell" function is to use knowledge to provide new products and services in order to consolidate customers and market and to create competitive advantages. Beckman's (1997) knowledge management process adopted by this paper with modified presentation is shown in Figure 1 below.

2.3. The source of knowledge

After the survey of literature, this paper puts forward four sources of knowledge: artifacts, practices, regulation and system and networking. Artifacts are man-made creations such as products, architectures, blueprints, legal sentences and documents. Practices refer to how to accomplish a task. Brown (1991) indicates that knowledge does not necessarily exist merely in blueprints and technologies that are directly related to research and development of new products; local innovation that directly improves how people work onsite is also the creation of knowledge. Therefore, he continues, the opportunities for knowledge creation exist at all levels of a business and in its external customers. Nonaka (1991) also indicates that it is an important issue for corporate knowledge management to make tacit knowledge, such as how experts work, explicit and systematic through socialization, articulation, combination and internalization. Regulation and system, such as work procedures and formal and informal regulations, are the implementation of certain practices throughout a business and the codification of them if necessary and possible. Networking is regarded as a referring and problem-solving system composed of human networks. For instance, to know more about a problem related to production, one can ask people in the production department directly. One can also ask one's colleagues and they may refer him or her to the person they believe to know the problem best. Davenport and Prusak (1998) stress strongly that the key to successful problem-solving in a business is whether or not there is effective networking. Mintzberg (1973) also points out that managers spend more than 40% of their time on networking and communication; they acquire most of their important information through word of mouth, hearsay and gossip because they believe these will become the facts of tomorrow.

3. THE RESEARCH DESIGN AND METHOD

3.1. The research design

Understanding the importance of building customer knowledge base through knowledge management, this paper aims to propose a conceptual framework for

exploring the relatedness among the three dimensions of resource provision, knowledge management process and customer knowledge performance as shown in Figure 2 below. To take a strategic view of knowledge management, the identification of the mission and vision and of core competences is crucial for a business because it can then know what kind of knowledge can improve its core competences, achieve its mission and, in the long run, actualize its vision. Case study research and a structured questionnaire survey are adopted as the research method. A kitchenware manufacturer and seller in Taiwan is the focus of the research whose management, through the analysis of its mission and

vision, have considered the improvement of its customer knowledge will enhance its core competences. The structured questionnaire survey is adopted to collect the concerned data for analyzing the expressed relatedness among the three dimensions of the proposed conceptual framework as reported by the respondents in the case company. The collected data will also be utilized to further analyze how customer knowledge can be enhanced through resource provision, knowledge management process and customer knowledge performance in this particular setting.



Figure 1: Corporate knowledge management process
Source: Adapted from Beckman (1997)

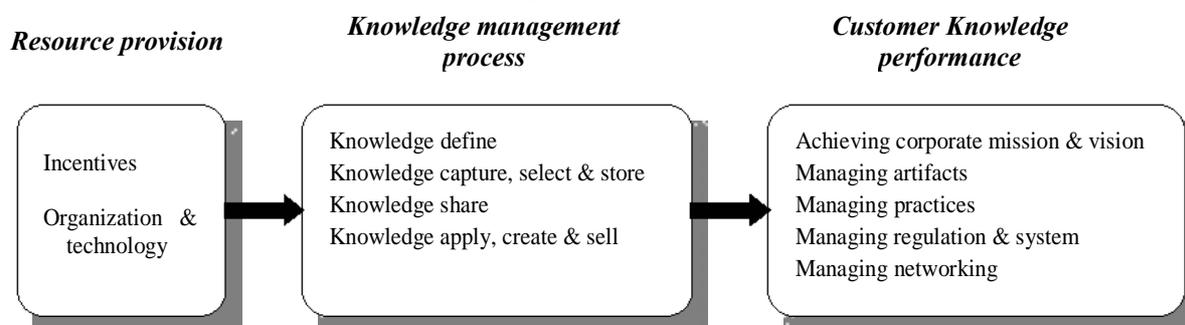


Figure 2: The proposed conceptual framework

For exploration of the dimension of resource provision, this paper proposes two major important resource inputs for successful corporate knowledge management including “Incentives” and “Organization and technology”, which have been mentioned above. For the exploration of the dimension of knowledge management process, this paper also adopts Beckman’s (1997) knowledge management process. The authors consulted senior managers of the case company and re-organized these eight steps into four steps, which would fit better their future planning for managing customer knowledge in their setting and could help them find out what step they should improve. These four steps were “Knowledge define”, “Knowledge capture, select and store”, “Knowledge share” and “Knowledge apply, create and sell”. For the exploration of the dimension of customer knowledge performance, this paper adopts the success of managing the four sources of knowledge - artifacts, practices, regulation and system and networking - as put forward by this paper above. Taking a strategic point of view of knowledge management process, this paper also adopts the success of achieving corporate mission and vision for the exploration of the dimension of customer knowledge performance. As mentioned above, clear corporate mission and vision enable people in a business to have clear objectives, sense of direction, priority, holistic consideration, coordinated actions and the basis of performance measures. They also set out the direction

of corporate knowledge management. Thus, it seems reasonable to assume that, under the clear guidance of corporate mission and vision, people who surely understand their corporate mission and vision can outperform in knowledge management process and in customer knowledge performance compared with people who do not. Based on the above, the objectives of the paper are to propose a framework which explores the relationship among the three dimensions of resources provision, knowledge management process and customer knowledge performance and show how the proposed framework can facilitate the enhancement of customer knowledge through resource provision and knowledge management process in this particular setting. The following hypotheses are also evolved:
 H1: Resource provision is positively associated with greater success of knowledge management process.
 H2: Knowledge management process is positively associated with greater customer knowledge performance.
 H3: Respondents who surely understand their corporate mission and vision achieve greater success of knowledge management process than respondents who do not.
 H4: Respondents who surely understand their corporate mission and vision achieve greater success of customer knowledge performance than respondents who do not.

3.2. Sample

As mentioned above, a kitchenware manufacturer and seller in Taiwan was selected for the survey setting. Founded in 1960s, the case company had stainless steel cookers as its major products and sold them in Taiwan, Japan, Northern America, Hong Kong, Australia and Malaysia. Highly emphasizing the importance of quality, the case company had been certified Japan's SG, U.S.A.'s UL and ISO 9001 and it had had the most advanced facilities for manufacturing stainless steel cookers in Asia. It had also devoted itself to continuous research and development to assure its customers of original taste, health and products' firmness. In the future, it had the vision of being the leading brand in the Chinese society. Despite its achievement, the case company had some serious problems such as how to differentiate itself from its competitors, increase market share and boost profits when the authors were invited for consultancy. Kitchenware market had been faced with severe competition worldwide. Building customer knowledge base through knowledge management provides for businesses an avenue for differentiating themselves from their competitors, increase market share and boost profits. A better understanding of customers made possible through knowledge management will enable customized products and services and sense of closeness and, subsequently, better customer relationships. Through the analysis of its mission and vision, the management of the case company considered the improvement of their customer knowledge being crucial for enhancing their core competences and long term competitiveness.

In terms of customer knowledge, the case company had invited its customers to a membership of its customer club. It provided for its members product warranty, services and direct mails. However, due to the lack of customer knowledge, the products' direct mails were delivered to all members without any segmentation. In the meantime, it also had little ideas about who its most profitable customers were. To know and serve its customer better and to enhance its customer knowledge through finding out for the case company in our study what resource provision and what step of knowledge management process it should improve were quite important under the circumstance. In our study, the unit of analysis was an employee in the case company. Questionnaires were sent to the employees in the various departments of the case company in its branches in Taiwan region. Respondents were asked to react using a five-point scale ranging from "Strongly disagree" to "Strongly agree". We got 54 responses from 54 questionnaires. The response rate was 100%.

3.3. Analysis

SPSS software package was adopted for data analysis in the paper. The data analysis methods were as follows.

3.3.1 Descriptive statistics

For individual variable or factor, descriptive statistics was adopted to describe the mean and standard deviation of data.

3.3.2 Reliability analysis

The value of Cronbach's alpha was adopted to examine the reliability of all constructs. For high reliability, the value of Cronbach's alpha should be greater than 0.7; reliability is low if its value is less than 0.3.

3.3.3 Analysis of regression

Analysis of regression was adopted by this paper to explore the relationship between the dimensions of resource provision and know management process and the relationship between those of knowledge management process and customer knowledge performance.

3.3.4 Analysis of variance

ANOVA was adopted by the study to examine the differences between respondents who surely understood their corporate mission and vision and respondents who did not regarding their success of knowledge management process and of customer knowledge performance.

4. RESULTS

4.1 Sample characteristics

Table 1 below summarizes the demographic characteristics of the respondents. Our sample was in the Taiwan region of the case company.

4.2 Instrument reliability

As mentioned above, the analysis of scale reliability was performed by computing and examining Cronbach's alpha coefficient. As shown below in Table 2, the values of Cronbach's alpha for most constructs were higher than 0.7 and were considered as reliable. Four components' reliabilities were between 0.5 and 0.6. They, however, still exceeded Nunally's (1967) suggested minimum reliability of 0.5 - 0.6 for instruments used in early stages of exploratory research. Due to the exploratory nature of the study, they were included in the subsequent data analysis activities.

4.3 Descriptive statistics of the three dimensions

4.3.1 The resource provision dimension

The authors had done interviews with some managers and employees in the case company after the questionnaire survey to verify the findings of the study. They agreed upon the findings presented to them. The mean scores and standard deviations of the constructs are shown below in Table 3. In the dimension of "Resource provision", the mean scores of the constructs were more than 3 and between 3.75 and 3.80. The respondents

reported moderate positive attitude towards the resource provision for knowledge management process in the case company.

4.3.2 The knowledge management process dimension

In the dimension of "Knowledge management process", the mean scores of the constructs were more than 3 and between 3.36 and 3.67. The respondents reported moderate positive attitude towards the success of knowledge management process in the case company. Among the four steps of knowledge management process, the respondents demonstrated the greatest reported success in the step of "Knowledge define" and the least reported success in the step of "Knowledge apply, create and sell" in the case company.

4.3.3 The customer knowledge performance dimension

In the dimension of "Customer knowledge performance", the mean scores of the constructs were more than 3 and between 3.29 and 3.86. The respondents reported moderate positive attitude towards the success of customer knowledge performance in the case company. Among the constructs of customer knowledge performance, the respondents demonstrated the greatest reported success in "Achieving corporate mission and vision" and the least reported success in "Managing networking" in the case company.

4.4 Analysis of regression

4.4.1 The association of the dimension of "Resource provision" with that of "Knowledge management process"

As shown below in Table 4, all the estimated regression coefficients were greater than zero and the component of "Organization and technology" was positively and statistically significantly associated with the success of four steps of knowledge management process. In the case company, however, our study did not provide empirical support for the association of "Incentives" with any of knowledge management process, which was quite contradictory against the claims on the importance of incentives for knowledge management in the literature. Nonetheless, in this particular setting, the case company had just decided to manage its customer knowledge systematically. All the steps of knowledge management process in the case company were voluntarily adopted by people rather done for incentives provided by the case company. Under this circumstance, the provision of delegated organization and information technology seemed more helpful for the success of knowledge management process. In the long run after knowledge management process is carried out systematically at corporate level in the case company, the component of "Incentives" might show stronger association with the success of knowledge management process as suggested by the literature. Interviews with managers and employees of the case company after the questionnaire survey confirmed this argument.

As also shown below in Table 4, "Resource provision" explained 40.7% of the variance of the step of "Knowledge define"; 59.9% of the variance of the step of "Knowledge capture, select and store"; 36.4% of the variance of the step of "Knowledge share"; and 36.5 % of the variance of the step of "Knowledge apply, create and sell". Overall, the amount of variance explained by the proposed framework appeared reasonable.

4.4.2 The association of the dimension of knowledge management process with that of customer knowledge performance

As shown below in Table 5, most of the estimated regression coefficients were greater than zero and the step of "Knowledge apply, create and sell" was positively and statistically significantly associated with the success of five customer knowledge performance components. That is, in the case company, "Knowledge apply, create and sell" made great contribution on the success of achieving corporate mission and vision and managing artifacts, practice, regulation and system and networking. On the other hand, however, the step of "Knowledge define" was only helpful for achieving the corporate mission and vision and the step of "Knowledge share" made considerable contribution only on managing artifacts. For the step of "Knowledge acquire, select and store", three of the five estimated regression coefficients were negative with two of the three at 5% significance level indicating that the step was even harmful for customer knowledge performance in the case company. The strong negative association implies that people who reported to use less knowledge acquired, selected and stored by the case company, achieved more success on managing artifacts and networking. This might be because the acquired, selected and stored knowledge regarding artifacts and networking was outdated, mistaken, incomplete or inappropriately documented. Interviews with managers and employees of the case company after the questionnaire survey confirmed this argument.

The step of "Knowledge define" was helpful for achieving the corporate mission and vision. The strong positive association implies that people who reported to accepted more knowledge defined by the case company, did better on achieving corporate mission and vision. This might be because the case company depicted a clear picture of its mission and vision and core competences and defined what sort of knowledge might help the enhancement of core competences and achievement of mission and vision. However, our study did not provide empirical support for the association of "Knowledge define" with the managing of artifacts, practices, regulation and system and networking. As mentioned above, in this particular setting, the case company had just decided to manage its customer knowledge systematically. The systematic definition for specific knowledge sources had not been implemented at the corporate level. Thus, at this stage, "Knowledge define" did not associate with the success of managing of

knowledge sources in the case company. The step of "Knowledge share" made considerable contribution on managing artifacts. The strong positive association implies that people who reported to use more knowledge shared within the case company, managed artifacts better. This might be because artifacts could be attained more easily through the step of "Knowledge share" within the case company compared with practices, regulation and system and networking due to their tacit nature. Our study did not provide empirical support for the association of "Knowledge share" with achieving corporate mission and vision despite that people who reported to accepted more knowledge defined by the case company, did better on achieving corporate mission and vision as mentioned above. This has an interesting implication that convincing people to accept the mission, vision and core competences defined by the case company, rather than circulating unconvincing mission, vision and core competences through the step of "Knowledge share" could be more helpful for achieving corporate mission and vision. Interviews with managers and employees of the case company after the questionnaire survey confirmed the arguments mentioned above.

As also shown below in Table 5, "Knowledge management process" explained 56.2% of the variance of the success of "Achieving corporate mission and vision"; 54.5% of the variance of the success of "Managing artifacts"; 61.5% of the variance of the success of "Managing practices"; 64.1% of the variance of the success of "Managing regulation and system" and 56.8% of the variance of the success of "Managing networking". Overall, the amount of variance explained by the proposed framework appeared high.

4.5 Analysis of variance

4.5.1 *The difference in the dimension of "Knowledge management process"*

As shown below in Table 6, for four of the four knowledge management process, respondents who surely understood their corporate mission and vision reported statistically significant higher scores than respondents who did not. Therefore, in the case company, respondents who surely understood their corporate mission and vision achieved greater success of "Knowledge define", "Knowledge capture, select and store", "Knowledge share" and "Knowledge apply, create and sell" compared with respondents who did not.

4.5.2 *The difference in the dimension of "Customer knowledge performance"*

As shown below in Table 7, for four of the five knowledge management performance measures, respondents who surely understood their corporate mission and vision reported statistically significant higher scores than respondents who did not. Therefore, in the case company, respondents who surely understood

their corporate mission and vision did better for "Achieving corporate mission and vision", "Managing artifacts", "Managing practice" and "Managing regulation and system" compared with respondents who did not.

5. CONCLUDING REMARKS

Under severe global competition, building customer knowledge base through knowledge management provides firms with strategic weapon to differentiate themselves from their competitors and to retain and strengthen competitiveness in the era of knowledge economics. Thus, this paper proposes a conceptual framework for exploring the relatedness of the three dimensions of resource provision, knowledge management process and customer knowledge performance and exploring how customer knowledge can be enhanced through resource provision and knowledge management. A case study has been provided for demonstrating and explaining how it operates. The proposed framework emphasizes the link of knowledge management to corporate mission and vision and that they set out the direction of corporate knowledge management. This study has taken a kitchenware manufacturer and seller in Taiwan as an example to perform a field study to collect the concerned data for exploring the expressed relatedness among the three dimensions of the proposed framework as reported by the respondents in the case company. As hypothesized, we found in the case company that resource provision was positively associated with greater success of knowledge management process and that knowledge management process was positively associated with greater customer knowledge performance. Also as hypothesized, we found in the case company that people who surely understood their corporate mission and vision achieved greater success of knowledge management process and of customer knowledge performance than people who did not. This study has taken as well the case company as an example to show how to enhance its customer knowledge through finding out what resource provision and what step of knowledge management process it should improve. This study has also demonstrated the link between corporate mission and vision and knowledge management as people who surely understand their corporate mission and vision can outperform in knowledge management process and in customer knowledge performance compared with people who do not.

ACKNOWLEDGEMENT

This research was supported by the National Science Council of Taiwan (grant NSC 92-2416-H-390-004)

Due to the space limitation, the full paper including tables, the discussion of implications and references is available upon request from Hwan-Yann Su.

Design and Development of an E-Learning Management System

Chee Chern Lim¹, Man Hing Yu¹, Jesse J. Jin¹, Vera Chung²

¹ School of Design, Communication and I.T., University of Newcastle, NSW 2308, AUSTRALIA

² School of Information Technologies, University of Sydney, NSW 2006, AUSTRALIA

{chern, zenoy, jesse, vchung}@it.usyd.edu.au

ABSTRACT

The trend of e-learning technologies is expanding fast. Web-based learning environments are becoming very common in the higher education institutions. Nowadays e-learning management systems are very popular. Many universities throughout the world deliver educational programs via the Internet. Developments of e-learning systems are generating great impact in the field of education services to improve the teaching and learning process, and overcome geographical distance. In recent years, various kinds of Internet technologies have become available for developers to implement such e-learning system that provide an e-learning gateway on the Internet. The rapid advancements in information and communication technologies, especially the networking and multimedia, have led to the development of many advanced e-learning systems these days. A user-friendly interface and a sophisticated data model are the essential design consideration to make the e-learning system easy-to-use for the instructors and learners. The need for such architecture is critical for designing the system and standards development. The system is developed under Computer Supported Cooperative Work framework and web portal technology. The system integrates all the critical and valuable communication tools that effectively improve the collaboration in an e-learning environment.

Keywords: e-learning, distance education, online education, virtual class

1. INTRODUCTION

Since the Internet emerged as a global information network, learning materials and information are available to any Internet user from any place at any time. For quite some times, people in education have been tremendously interested in using the facilities provided by the technologies of information and communication (TIC) [1]. In the recent years, the rapid advancements in technology and telecommunications, especially the Internet, have led to an explosive growth of web-based courses. The World Wide Web (WWW) is emerging as the most popular medium for delivering instruction and an increasing number of instructors are integrating various features of the Web in novel ways to provide exciting learning opportunities for their students. In addition, the use of Internet for education purposes has revolutionized distance learning. People have been tremendously interested in delivering online teaching and distance learning over the Internet. WebCT and Blackboard are the two of the most comprehensive systems widely used in most higher education institutions [2][3]. Course management system is a general term for server capabilities enabling online teaching functions, including presentation of contents, discussion groups, class management and related services. WebCT and Blackboard are the two e-learning systems which provide the most comprehensive tools and capabilities to support a sophisticated Internet-based learning environment. This paper introduces a new web-based toolkit Cyber Campus for instructors to deliver database-driven web systems in e-learning management.

This paper will provide an overview of the system structure in the Cyber Campus. Next it will focus on issues relating to the development of the communication and assessment modules in the system. These issues include some of the enabling technologies used in the development. After that, the performance of the system will be examined and compared with other e-learning system. Finally, this paper will be concluded with a summary of the system.

2. CYBER CAMPUS E-LEARNING SYSTEM

Cyber Campus is an all-in-one web-based e-learning system that integrates components including student management, course management, assessments, classroom allocation and communication into one complete package solution. The system is designed by using the framework based on Web Content Component Model (WCCM), which is a suitable model of web application that supports maintenance of content oriented web application [4]. In addition, the components will be developed using portal technology. An e-learning web portal serves as an integrated gateway in the online education centre website and provides users (instructors, learners and administrators) with a single point of access to the e-learning services [5]. Cyber Campus is implemented in PHP, MySQL and Java technology [6][7][8]. Figure 1 shows that the system provides users with a single access point to the e-learning portals.

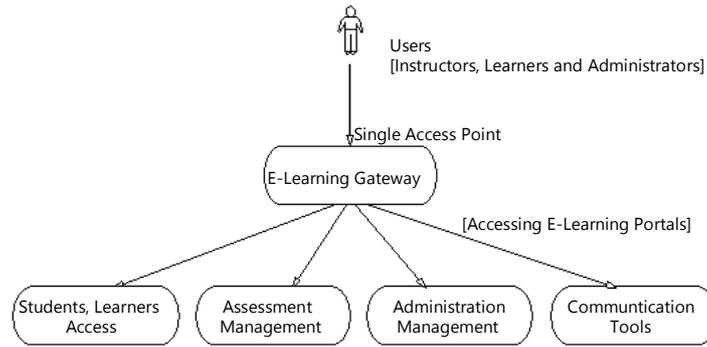


Figure 1. E-Learning Portal Gateway

Figure 2 shows the interaction between users under an user-friendly administrating, teaching and learning e-learning environment. Cyber Campus integrates web environment. It has provided administrators with a technology with database system and provides a centralized control in the entire system.

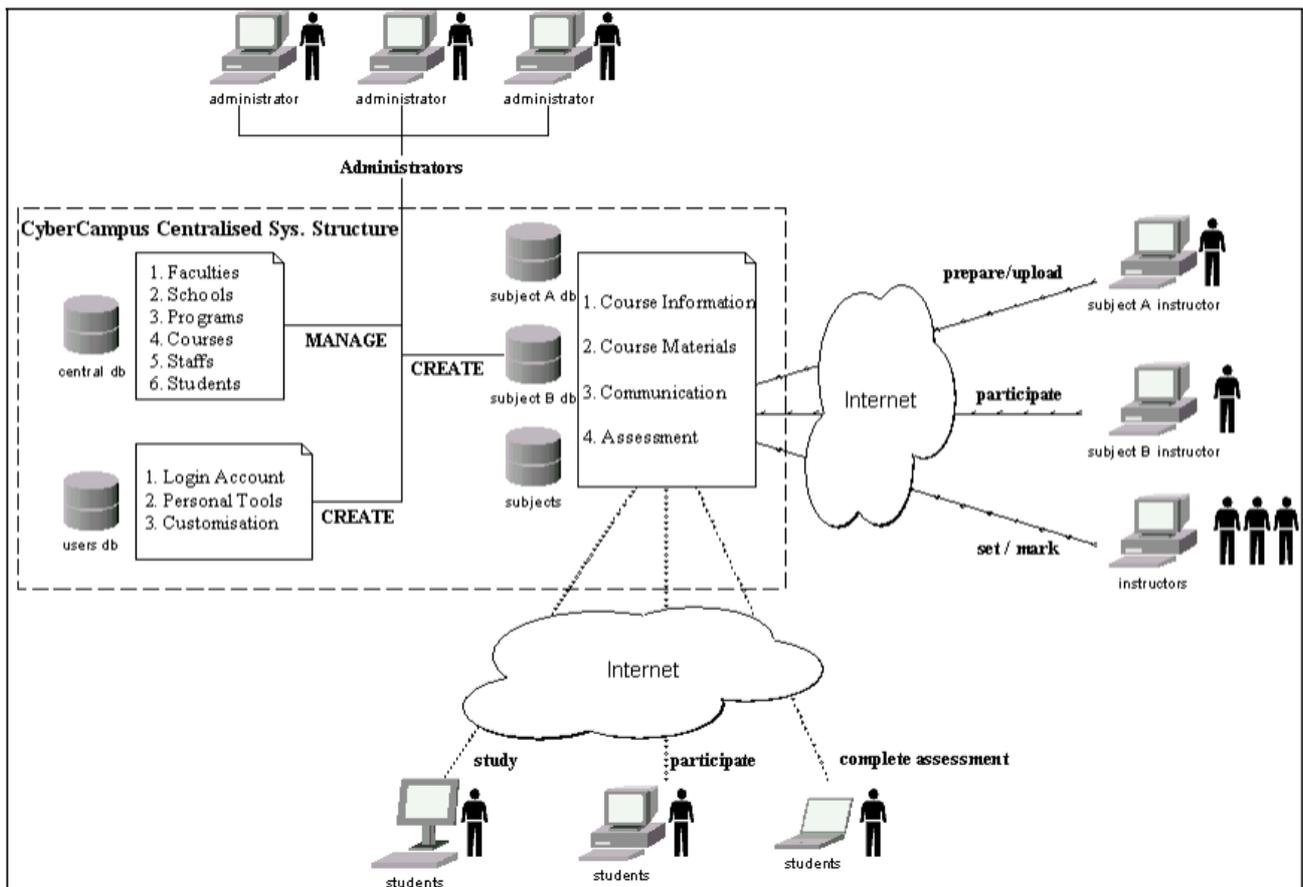


Figure 2. Cyber Campus e-Learning Environment

In addition, it includes many useful online education tools for the instructors and learners. The administration management module contains a set of functions, which help the administrators to manage the information resources in an easier and more efficient way.

3. SYNCHRONOUS CHANNELS COMMUNICATION

Cyber Campus is designed with a very comprehensive set of communication component on top of the core course management framework. This communication module is designed and implemented according to the concept of CSCW (Computer Supported Cooperative Work) [9]. Individually, the communication tools that Cyber Campus is designed of have many competitors that are currently in use by the corporate community. The synchronous collaboration component features the following sub modules:

3.1 E-Chat

The E-Chat module is written entirely with core Java and Java applet. The applet had to be compatible with Java 1.1.5 so as to work with Microsoft's default Java VM plug-in for Internet Explorer. The objective of this design is to ensure that the applet can be run from the most common web browser (i.e. Internet Explorer or Netscape Communicator). The online chat system is designed to be easily extendible and configurable no matter what sort of system it is deployed on, with minimal dependencies. The server includes a multithreaded object class, which spawns threads when users connect through the Java applet embedded in the dynamic html page. A maximum client limit is

imposed so as to hinder Denial of Service attacks. This design enables the server to handle multiple clients without heavy demand on the system.

The client's Java applet is designed to run from any web browsers. It is designed in a highly object-oriented behaviour, with the separation between the different design patterns of the back-end functions and the front-end graphical user interface. In addition, the client and the server communicate via TCP socket, which provides reliability and error correction in the network environment. Every e-chat session is saved in the database system. The user may, at a later stage, view the session information and the chat contents in each session. Instructors and learners can review any previous online chat discussion any time. Figure 3 shows the E-Chat module and its sub components.

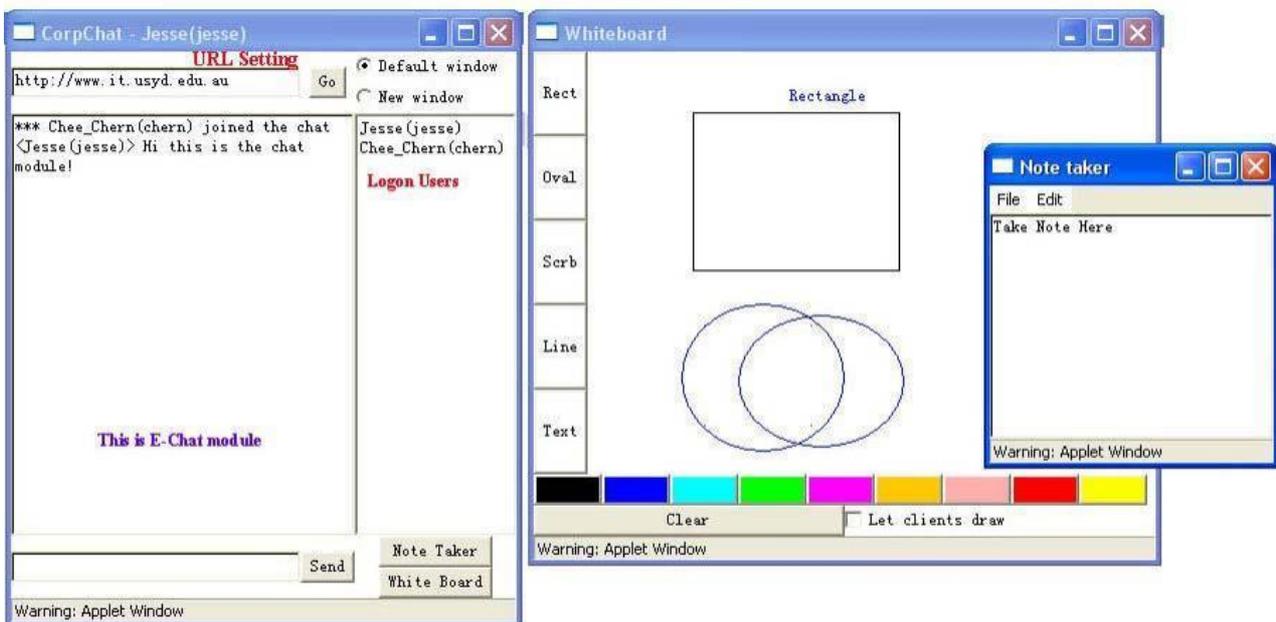


Figure 3. E-Chat Module And Its Components

3.2 Peer-to-peer File Transfer

In the e-chat module, the system also provides a 'File Transfer' function. This function allows active user in chat session to send file to other user in the same session. This function helps to improve the collaboration between the system users. In most existing system, users can only exchange files through email. Thus, any user without email is not able to communicate with other users. Cyber Campus has provided another alternative way for users to communicate in a much efficient way.

3.3 Electronic Whiteboard

In order to further improve the collaboration for students and tutors in their e-learning environment, the idea of having an electronic whiteboard is implemented

in the Cyber Campus. The whiteboard allows students to perform drawing on the Internet. Users can share ideas and communicate in a graphical ways. Users can enter text; draw lines, rectangles, scribbling and ovals of any colour. The whiteboard is designed to have the option of saving the drawing and opening previous saved drawing or other standard image files. Again, this improves the communication and collaborative group work between the users. This component is similar to the E-Chat module, it is implemented purely in Java, and thus users can operate from any standard web browser on any platform.

3.4 Synchronized Group Web Browsing

Synchronized group web browsing is the next feature in the communication tools. With the combination of chat and synchronized group web browsing, users can be notified automatically when anyone joins or leaves

the session. The system is just like a virtual classroom session, student 'Chee' joined instructor in the chat and the user in the same session can interact with each session. The instructor can set an URL address in the other. In addition, instructor can direct active users' chat applet. When an URL is set, the instructor and browsers to the designated page by setting an URL every student in same session will have browsers address. Unlike other solutions, by utilizing Java displaying the contents from the preset specific URL. technology, this module can be directly integrated in the Instructor can use this tool to deliver teaching on the Cyber Campus without any additional effort. Figure 4 Internet. This achieves the basic requirements of a virtual classroom. Instructor 'Jesse' creates a new

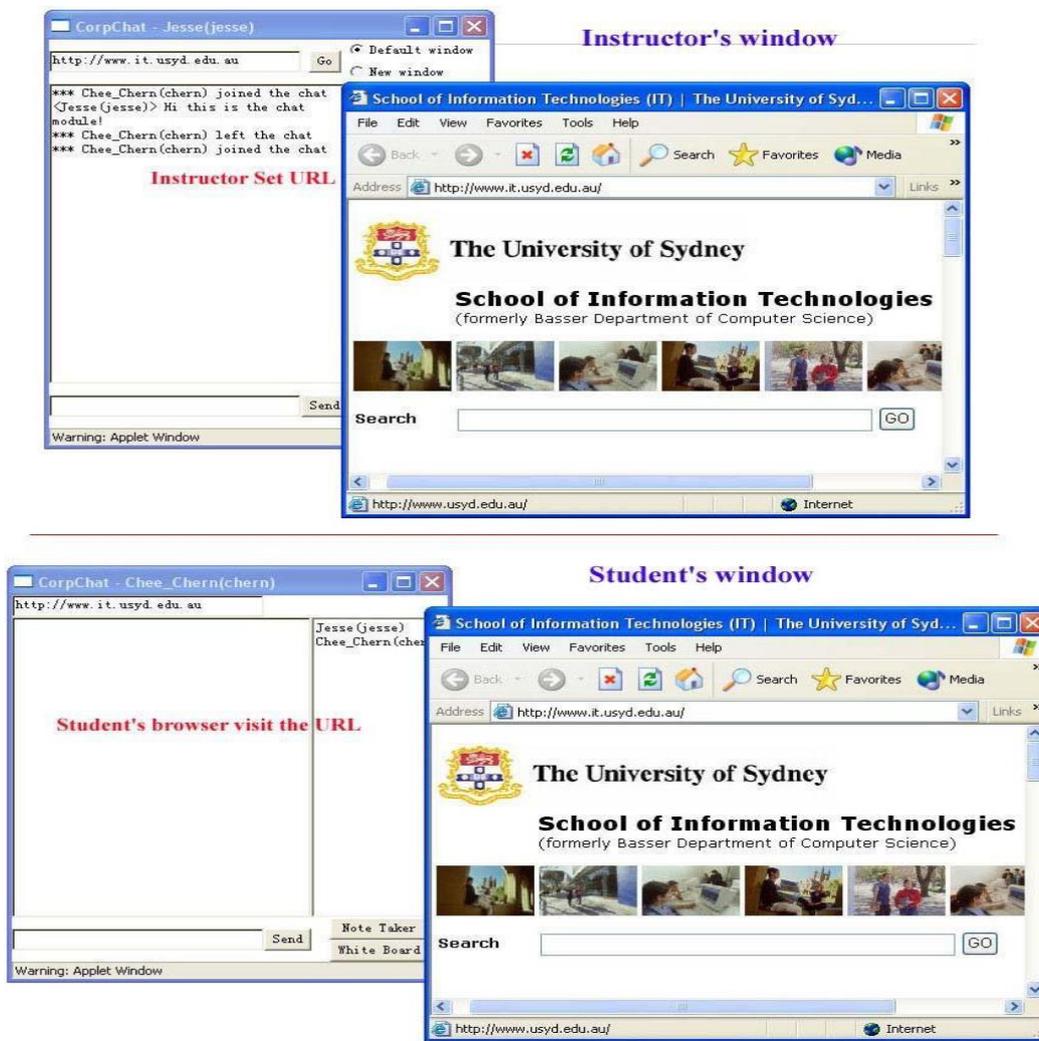


Figure 4. Synchronized Group Web Browsing Between Users

4. ASSESSMENT MANAGEMENT

4.1 Online Assessment

Assessment management is a very important module in any e-learning system. Cyber Campus has provided a wide range of options in setting up different kinds of assessments. The framework allows instructor to either create online assessment questions or upload project's related documents. In addition, the system allows instructor to create submission gateway and perform auto-marking on the online assessments. The following sections will further explain the design and functionalities in the assessment module.

In Cyber Campus, assessments are identified into two categories in term of the assessment nature: creating questions online or uploading the files. Instructor can easily setup assessment questions through a user-friendly interface. The system provides tools to construct questions including the following question format: multiple choice, true or false, short answer or essay-type questions. In addition, in the process of creating or editing assessment details, instructors can configure additional settings such as assessment's online

status, date, time and duration, time limit on each question, solution and explanation. Once the assessment and questions are properly setup, students can login to the system and participate in the test. All students' answers will be recorded for later use in the assessment marking.

4.2 Online Submission

For non-online assessment or non-hardcopy submission, instructors can setup gateways to receive softcopy or electronic submissions of assignment from the students through the Internet. For each submission gateway, instructors can set some restrictions such as submission's due date, allowed number of submitted files, acceptable file extensions, late penalty and its online status. Furthermore, instructor can define each own marking scheme and use it as marking sheet in marking the submissions.

4.3 Assessment Marking

This section can be further divided into four modules:

1. Auto-marking in online assessment.
2. Marking sheet on students' submissions.
3. Enter mark for traditional assessment.
4. Normal distribution on mark scaling.

For the online assessment with multiple choice or true/false questions, the system can perform auto-marking on the students' answers. However, for short answer and essay-type question, instructor is still required to review the answer and enter the mark manually.

From the submission management, instructor can view and retrieve students' submissions and mark the submissions with pre-define marking scheme. Late penalty will automatically be calculated upon the submission received from the student. Instructor can download submission's files and view history record of each student's submission.

As for traditional assessment, Camber Campus has provided an additional interface for instructors to manually enter the result into the system. Finally, instructor can perform mark scaling with function calculating the normal distribution. With graphical presentation of the results, instructor can easily identify the marks distribution of students' results.

5. SYSTEM PERFORMANCE

Cyber Campus e-learning system is successfully deployed in the School of Information Technologies, University of Sydney. One subject is selected to participate in the beta version of the system. The system is installed on a Window platform machine running on an Apache server. The system is used to serve as an e-learning system that provides instructor and learners with the full system functionalities

6. CONCLUSION

This paper introduced an e-learning system that integrates students management, course contents management and users collaboration into an all-in-one web-based application. The framework is carefully designed according to the model of WCCM, and the entire system will be implemented using portal technology [4][5]. In addition, the communication component is designed according to concept of CSCW [9]. The web-based multimedia e-learning system provides a centralised control in the database. CyberCampus is platform independent. It can be installed on any server and users can access the system through Internet/Intranet using standard web browser regardless of the client's operating system. The efficiency of collaboration and consultation is improved by providing an online chat session and simultaneous group web browsing environment that user could aware the presence of other participants. In addition, the system supports peer-to-peer connection for efficient files transmission between logon users. However, the system is still currently under constant improvement in the development phase. The latest version of whiteboard is still unable to save the drawings and open imaging file. Overall, Cyber Campus satisfies the requirements of a virtual classroom. The system will be the ideal system for any size of education organisation and it has proven its stability and abilities in its beta version.

ACKNOWLEDGEMENT

ICEB2004 is supported by the National Natural Science Foundation of China.

REFERENCES

- [1] Lelouche, R. and Page, M, "WebCT A Tool for Proselytism!", Proceedings of 2002 IEEE International Conference on Computers in Education
- [2] WecCT.com (<http://www.webct.com>)
- [3] Blackboard Inc (<http://www.blackboard.com>)
- [4] Li, Qingshan, "Study of A Content Oriented Web Architectural Model", IEEE 2001.
- [5] Wege, Christian, "Portal Server Technology", IEEE Internet Computing, May - June 2002.
- [6] PHP: Hypertext Preprocessor. (<http://www.php.net>)
- [7] MySQL: The World's Most Popular Open Source Database. (<http://www.mysql.com>)

- [8] Sun Microsystems (<http://www.sun.com>)
- [9] Bentley, R. & Horstmann, T. & Trevor, J., "The World Wide Web as enabling technology fro CSCW: The case of BSCW", Computer Supported Cooperative Work: Special issue on CSCW and the Web, Vol. 6. 1997

Effective E-Teaching: The Case of Hong Kong

Candy K. Y. Ho¹, Ivy Chan²

¹ Department of Marketing, The Chinese University of Hong Kong, Hong Kong, China

² Department of Decision Sciences and Managerial Economics,
The Chinese University of Hong Kong, Hong Kong, China
{kyho, ivychan}@baf.msmail.cuhk.edu.hk

ABSTRACT

Despite the heavy promotion by the local government, the plentiful provision of IT educational tools, and the sufficient IT skills of prospective users, the adoption of IT in classroom education by primary and secondary school teachers in Hong Kong is still minor. By focusing on the use of IT educational Websites, which are freely available for teachers, this paper attempts to investigate the factors that influence the adoption of e-teaching in classrooms among primary and secondary school teachers in Hong Kong. With reference to the paradigms of information technology, cognitive thinking and organizational learning, a research model, together with a preliminary outline of the operationalization of the measures and data collection method for model testing, are presented in this paper. It is believed that the findings of the study proposed here could contribute to a more in-depth understanding on the adoption of IT tools by teachers, which is crucial for software developers to excel in this potentially enormous industry.

Keywords: e-teaching, primary and secondary education, IT adoption, educational Website

1. INTRODUCTION

The use of Information Technology (IT) in primary and secondary education has recently been receiving much attention in Asia, primarily due to the boom in using the Internet in the region as well as the potential in exploiting this medium to enhance interactive communication and collaborative learning (Kumar, 1996; Wong, 2000). In Hong Kong, there are various educational parties or software developers providing IT educational tools (e.g., animated movies and interactive games) for primary and secondary school teachers free of charge, with dedication to host them through the Internet for easy and extensive access (Table 1). Conceivably, the propagation of those materials could enable teachers to leverage multiple ways of classroom communication, information sharing, and learning effectiveness (Puntambekar et al., 1997; Wong, 2000). However the usage of IT (specifically adopted as a mean for effective

teaching) has not received as much interests and enthusiasm from teachers as expected.

In regard to the evolving growth of e-based education, we consider that a systemic inquiry of the disparity between the devoted propagation and actual adoption is crucial. Specifically, as teachers are the initiators of classroom e-education, this paper investigates the factors that influence the adoption of e-teaching among primary and secondary school teachers in Hong Kong. Section 2 describes the recent trends in IT development and usage in the educational sector of Hong Kong, which give rise to the motivation of a study proposed in this paper. Section 3 presents the highlights of e-teaching as well as e-learning in previous studies. Section 4 depicts a research model together with a preliminary outline of the operationalization of the measures and data collection method for model testing. Section 5 proposes the agenda for the study. Finally, a conclusion is given in Section 6.

Table 1. IT Educational Tools' Developers and Educational Websites

<u>IT educational tools developer</u>	<u>Examples of IT educational Website</u>
Statutory body e.g. The Educational and Manpower Bureau http://www.emb.gov.hk	http://resources.ed.gov.hk/funcomputer/
Non-profit making educational organizations e.g. HKEdcity http://www.hkedcity.net	http://ihouse.hkedcity.net/~ma3711/
Commercial Organizations e.g. Textbook publishers	http://www.timeskids.com.hk (Movie and game section)
Primary and secondary school teachers	http://hk.geocities.com/math1addminus/a1.htm

2. EDUCATION AND IT IN HONG KONG

As mentioned above, it is mused that e-teaching is active in Hong Kong. The majority of the education tools are offered for use free of charge, and the government's 5-year strategy on IT in education launched in 1998 has substantially devoted financial resources, manpower and infrastructures in promoting the use of IT in primary and secondary education. Despite the availability of favorable condition of e-teaching, it is found that there are unmet promises and thwarted performance in regard to the low usage and slow diffusion. As of year 2000, about half of the primary and secondary school teachers still spent less than 5% of their teaching time in using IT (Education and Manpower Bureau, 2001). Besides, those adopted IT in their teaching tended to use primitive yet ineffective tools, which could hardly facilitate two-way communication or enhance learning effectiveness.

At present, little has been known about the reasons of the relatively low adoption rate of IT tools in classroom settings. As teachers are the initiators and reference stance of IT adoption in classroom learning, our study focus on the examination of the genuine factors that facilitate or inhibit e-teaching adoption in primary and secondary schools in Hong Kong. We ultimately identified studies related to IT adoption, learning, knowledge management and cognitive thinking are insightful and useful as the research paradigms for the proposed model.

3. E-TEACHING AND E-LEARNING

Educational reports seem supporting that time is a crucial factor to affect teachers' e-teaching behavior. For those teachers adopted electronic means in teaching mused that they have to spend a lot of time in taking up the technology, prior to the integrative use of with teaching content. Other teachers dropped out from e-teaching asserted that their time for e-teaching were distracted with the increasing load in other school work. Other non-adopters claimed that they may not draw much benefits from e-teaching. Other inhibiting factors include: lack of management support and oversight, lack of motivation to transit from traditional classroom teaching to e-teaching, incapable to cope with technology, lack of measurement on teaching effectiveness from adopting e-teaching.

Despite these hindrance, the benefits of adopting IT tools in learning has been ascertained. There have been studies that address the significance for e-learning in enhancing the organizational performance (e.g., Benbunan-Fich 2002, Frankola 2001, Honey 2001), types of learning (Jansen et al. 2002, Lytras et al. 2003, Taylor 2002) or success factors (e.g., Hiltz and Turoff 2002, Kerrin and Winfield 2001). Yet, only a few of them recognize learning as a process that transited from conventional face-to-face mode (teacher as dominant

player) to e-teaching (teacher in complement with technology) and e-learning (students as dominant player) (Crocetti 2002). Given the infant stage of technology deployment in the educational sector of Hong Kong, this study aims to examine the behaviors of teachers who are primary user groups, as such we can capitalize the values of technology in classrooms (e-teaching) and propagate to students for well-rounded learning process (e-learning).

4. PROPOSED MODEL

The innovative way for learning with IT as the medium of teaching encompasses a complex blend of socio-technological factors that influence the usage the technologies involved. A review of prior literature and empirical studies in regard to the paradigms of information technology, cognitive thinking and organizational learning was conducted. Several key notions were emerged that enable us to devise a proposed model for further study. It was found that the Technology Acceptance Model (TAM) (Davis 1989) is seminal as a starting point to investigate plausible elements that affect (either as enablers or inhibitors) the teachers to adopt IT in their teaching practice. In view of the intricate nature of individuals and their intersubjective interpretations towards the complex environment (Kettanurak et al. 2001), it is suggested that cognitive aspect is recognized as the precondition to govern one's behaviors and actions. In particular, the cognitive styles or personal traits have been attended and researched to explain the desirable outcomes undertaken by human (Akbar 2003, Choi and Lee 2003), which implies the relevance in the current studies. Exploring the learning cycle and knowledge management, it is deemed that the effective use of technology by teachers can be collated with the degree of diffusion of innovation (Rogers, 1995). More subtly, the diffusion can be extended with the depth and breadth of diffusion (Malhotra and Galletta 1999).

The proposed model is presented (Figure 1). The perceived usefulness and perceived ease of use are independent variables that affect the actual use (including infusion and diffusion encompassed within dotted line). The technology usage is mediated by the attitude presented by the users. Between the independent variables and mediating factors, it is proposed that the human traits as innovativeness, and the environmental context as social influence are moderators that influence the associations.

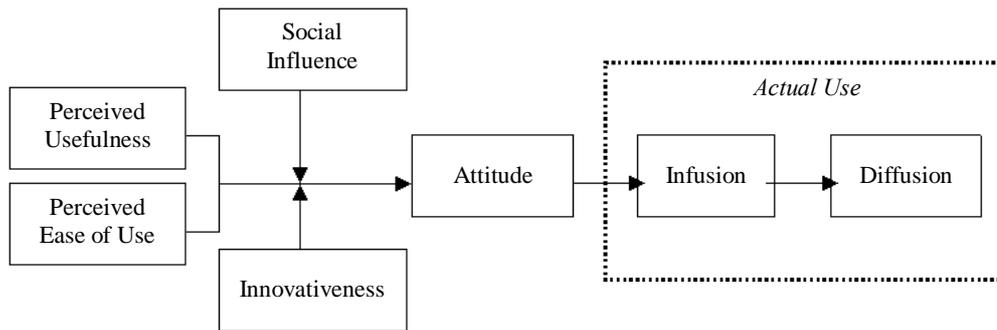


Figure 1. Proposed Model of E-Teaching in Hong Kong Education

Table 2. Operationalization of Measures – an Example

Constructs	Sample Measures in E-Teaching Context *	Reference
Perceived Usefulness Defined as the user's (i.e. teacher in this study) subjective belief that using a specific application system will increase his job performance.	Using X in my teaching would enable me to accomplish tasks more quickly.	Davis (1989) Chau and Hu (2001)
Perceived Ease of Use Defined as the extent of user expects that using a target system to be free of effort.	Learning to operate X would be easy for me.	Davis 1989 (Chau and Hu 2001)
Social Influence Defined as the extent of user's affection change or commitment to technology due to external force, for example peer group pressure or government mandates.	Colleagues in my school think we should avoid to use X. Government believes teachers should be enthusiastic to use X.	Malhotra and Galletta (1999) Lewis et al. (2003)
Innovativeness Defined as the degree user's inclination/personal traits towards doing things better (adaptive) or doing things differently (innovative).	Readily agrees with the team at work (Adaptor). Prefers changes to occur gradually (Innovator).	Kirton (1976) Garfield et al. (2001)
Attitude Defined as the user's desirability of his use of system.	Using X in my teaching is a good idea.	Taylor and Todd (1995)
Infusion Defined as the user's state in making use of a target system in a full and complete manner.	I have a thorough knowledge using X in my teaching. I can integrate the target system X in different teaching subjects.	Kishore and McLean (1998) Carter et al. (1999)
Diffusion Defined as the user's state in persuading other in the use of a target system.	I recommend X to my colleagues for effective teaching. I am committed to let my colleagues to explore the potency of X in teaching.	Kishore and McLean (1998) Carter et al. (1999) Lewis and Orton (2000)

* X – denotes as a selected educational website

As illustrated in Figure 1, there are seven constructs entailed. The description of each construct is detailed in Table 2, with their examples of measurement item and reference source.

5. RESEARCH AGENDA

Understanding the technology adoption in Hong Kong education is significant to the policy makers and educators as it directs the ways to advance knowledge and competences of students, and competitiveness of people in the long run. Also, an understanding of the

factors that affect the adoption and diffusion of IT educational tools among teachers enables software developers to develop and promote their educational products in an effective and efficient manner. Therefore, the exploratory study needs further synthesis and empirical validation. A seven-stage of research task is planned.

1. Adopt validated instrument/ measurement items (Table 2)
≥ Undertake minor wording changes to tailor them to the target context of education

2. Conducted Focus Group study to solicit qualitative data and opinion
 - ≥ Expose other unattended attributes of each construct
3. Synthesize and frame up results from stage (2)
 - ≥ Augment to the measure items in stage (1) and prepare initial questionnaire for pretest
4. Pretest with a sample of 50 teachers on their E-teaching behaviors
 - ≥ Include both adopters and non-adopters in the subject sample
5. Assess data for statistical generalization and refine the items
 - ≥ Base upon the validation on reliability and validity
 - ≥ Prepare the questionnaire for main survey
6. Collect data from the schools with IT adoption and without IT adoption in teaching
 - ≥ Seek help from government offices to institutionalize the survey
7. Compare results and generalize the features in each group
 - ≥ Draw implications on enhancement program on IT usage, for example culture, training and development and system design

6. CONCLUSION

The study proposed in this paper aims to explicate the underlying reasons for the (in)effective e-education phenomena in Hong Kong, specifically e-teaching. It is expected that the systemic approach of examination can shed lights on the entities for primary and secondary education, such as the government, management and teachers, in order to evaluate their planned initiatives, existing practices and possible remedies.

REFERENCES

- [1] Akbar, H., "Knowledge levels and their transformation: Towards the integration of knowledge creation and individual learning", *Journal of Management Studies*, Vol. 40, No. 8, pp1997-2021, 2003.
- [2] Benbunan-Fich, R., "Improving education and training with IT", *Communications of the ACM*, Vol. 45, No. 6, pp94-99, 2002.
- [3] Carter, P. E., R. Agarwal, V. Sambamurthy, "Organizational innovation with information technologies: The cycle of adoption, adaptation and use", *proceedings of the Americas Conference on Information Systems 1999*, pp. 585-587.
- [4] Chau, P.Y.K., P.J. H. Hu, "Information technology acceptance by professionals: A model comparison approach", *Decision Sciences*, Vol. 32, No. 4, pp699-719, 2001.
- [5] Choi, B., H. Lee, "An empirical investigation of KM styles and their effect on corporate performance", *Information & Management*, Vol. 40, No. 4, pp403-417, 2003.
- [6] Crocetti, C., "Corporate learning: A knowledge management perspective", *The Internet and Higher Education*, Vol. 4, No. 3-4, pp271-285, 2001.
- [7] Davis, F. D., "Perceived usefulness, perceived ease of use and user acceptance of information technology", *MIS Quarterly*, Vol. 13, No.3, pp319-340, 1989.
- [8] Education and Manpower Bureau, *Preliminary Study on Reviewing the Progress and Evaluating the Information Technology in Education (ITEd) Projects (December 2000 – August 2001)*, http://www.emb.org.hk/ited/Documents/ITEd_Report/FinalReport_v3.0_web.htm, accessed in July 2004.
- [9] Frnakola, K. "Why online learners drop out", *Workforce*, pp52-60, Oct 2001.
- [10] Garfield, M., N. Taylor, A. R. Dennis, J. W. Satzinger, "Research report: Modifying Paradigms – Individual differences, creativity technique, and exposure to ideas in group idea generation", *Information Systems Research*, Vol. 12, No. 3, pp322-333, 2001.
- [11] Hiltz, S. R., M. Turoff, "What makes learning effective?", *Communications of the ACM*, Vol. 45, No. 4, pp56-59, 2002.
- [12] Honey, P., "E-learning: A performance appraisal and some suggestions for improvement", *The Learning Organization*, Vol. 8, No. 5, pp200-202, 2001.
- [13] Jansen, W., H. M. van den Hooven, H. P. M. Jagers, G. C. A. Steenbakkens, "The added value of e-learning", *Proceedings of the InSITE 2002*, pp733-746.
- [14] Kettanurak, V., K. Ramamurthy, W. D. Haseman, "User attitude as a mediator of learning performance improvement in an interactive multimedia environment: An empirical investigation of the degree of interactivity and learning styles", *International Journal of Human-Computer Studies*, Vol. 54, No. 4, pp541-583, 2001.
- [15] Kirton, M., "Adaptors and innovators: A description and measure", *Journal of Applied Psychology*, Vol. 61, No. 5, pp622-629, 1976.
- [16] Kishore, R., E. R. McLean, "Diffusion and infusion: Two dimensions of success of adoption of IS innovations", *Proceedings of the Americas Conference on Information Systems 1998*, pp731-733.
- [17] Kumar, V. S., "Computer-supported collaborative learning: Issues for research", <http://www.cs.usask.ca/grads/vsk719/academic/890/project2/project2.html>, accessed in July 2004.

- [18] Lewis, N. J., P. Orton, "The five attributes of innovative e-learning", *Training and Development*, pp47-51, June 2000.
- [19] Lewis, W., R. Agarwal, V. Sambamurthy, "Sources of influence on beliefs about information technology use: An empirical study of knowledge workers", *MIS Quarterly*, Vol. 27, No. 4, pp657-678, 2003.
- [20] Lytras, M. D., A. A. Tsilira, M. G. Themistocleous, "Technology classification framework for e-learning purposes from a knowledge management perspective", *Proceedings of the Americas Conference on Information Systems 2003*, pp2573-2582.
- [21] Malhotra, Y., D. F. Galletta, "Extending the technology acceptance model to account for social influence: Theoretical bases and empirical validation", *Proceedings of the Thirty-second Hawaii International Conference on System Sciences*, 1999.
- [22] Puntambekar, S., K. Nagel, R. Hübscher, M. Guzdial, J. L. Kolodner, "Intragroup and intergroup: An exploration of learning with complementary collaboration tools", *Proceedings of Computer Supported Collaborative Learning Conference 1997*, pp207-215.
- [23] Rogers, E. M., *Diffusion of Innovations*, 4th Ed., New York: The Free Press, 1995.
- [24] Taylor, R. W., "Pros and cons of online learning – a faculty perspective", *Journal of European Industrial Training*, Vol. 26, No. 1, pp24-37, 2002.
- [25] Taylor, S., P. A. Todd, "Understanding information technology usage: A test of competing models", *Information Systems Research*, Vol. 6, pp144-176, 1995.
- [26] Winfield, I. J., M. Kerrin, "Catalyst for organizational learning", *The Learning Organization*, Vol. 1, No. 3, pp4-9, 1994.
- [27] 黃寶財, "互聯網在教育的應用 - 怎樣用 ? ", http://www.hkedcity.net/iclub_files/a/1/32/webpage/seminar/pcwong20010601.html, accessed in July 2004.

E-Learning Performance and Students' Results Case of a French Business School

Emmanuel Houzé¹, Régis Meissonier²

¹ CREGO - IAE – Montpellier II University – France; houze@iae.univ-montp2.fr

² CEROM – Montpellier Business School – France; rmeissonier@supco-montpellier.fr

ABSTRACT

E-learning experiments in higher education are becoming more recurrent. However, these experiments are seldom tangibly applied to an entire academic year group. Integrating *e-learning* into a pedagogical program implies performance analysis in terms of both students and teachers, but also from the institute's point of view. Due to the lack of Information Systems based research into *e-learning* performance modeling, the article propose an analysis mixing this research area with some findings in Education Sciences.

The first part of this article presents an analysis of the main scientific publications on which we have built our research model. The second part presents the initial findings of our on-going research project at Montpellier Business School (France). A comparison between traditional teaching and face-to-face teaching was carried out using the student marks in five different courses of study. The results show that the teachers' predisposition to adopting these new teaching techniques is not directly related to an improvement in the students' results. In other words, the paper is consistent with the need to avoid any techno-centered approach to on-line education.

In the same way, the article concludes that a measure of the e-learning performance must not be limited to the students' results alone. Indeed, the case studied puts forward that the legitimacy of an *e-learning* project can lie more in the satisfaction of being able to meet new strategic challenges through its development, than in simply improving an existing teaching tool.

Key-words: *E-learning*, Performance, Learning, Higher studies, Business School

1. INTRODUCTION

Given the current economic context, in which the two driving forces are the globalization of trade and the development of information technologies, the sector of higher education is undergoing a genuine transformation. The e-learning phenomenon represents one of the most highly developed computer-assisted approaches in this educational evolution.

The gloomy economic climate, since 2002, in which this "new cognitive economy" is taking shape, proves (yet again) that the exhilaration expounded by an innovative concept does not automatically go hand in hand with the anticipated, or even expected, effectiveness. Consequently, introducing *e-learning* into the education sector reveals some obscurantism about its own performance.

Few scientific works have been published on the subject compared to other areas of application of information systems. This indicates the lack of theoretical corpus sufficiently consensual to come to terms with this issue. The contingency of the cases studied or even the newness of the concept could certainly be cited as causes. But, such observations do not legitimate research projects flouting theories which, despite being inherent, can nevertheless be associated with it.

In Management Sciences, it can be considered that research into information systems involves the contribution of organizational and technological

mechanisms that bring media to knowledge and information exchange. Traditionally, an information system is assessed from different view points, depending on: its efficiency, its effectiveness and how players use it and are satisfied with it. But, specific *e-learning*-related criteria have not been clearly identified. This is why the assessment criteria can, to a large degree, be taken from those used in education sciences, a field in which a lot of research has gone into how to assess learners (using technology or not). Therefore, we have chosen to combine these two fields, information systems and education sciences, in order to propose an *e-learning* experimentation performance analysis grid. The first part of this article presents an analysis of the main scientific publications on which we have built our research model. The second part presents the initial findings of our on-going research project at the Montpellier Business School (France). A comparison between traditional teaching and face-to-face teaching was carried out using the student marks in five different courses of study. The results show that the teachers' predisposition to adopting these new teaching techniques is not directly related to an improvement in the students' results.

For the moment, the initial data collected only allows us to conduct a comparative analysis of the marks obtained by students having attended these on-line lessons and the marks gained by those who previously attended traditional classes for the same courses. These results should be interpreted in the broader context of learning processes and overall performance measuring. The tests

presented hereafter, on the influence levels of various characteristics and factors, will be the subject of future publications.

2. LITTERATURE ANALYSIS

Even if *e-learning*-based scientific publications in the educational field are less verbose than in other new technology fields of application, their combination makes it possible to envisage what the basis of an

analytical model - to gauge the performance of remote teaching approaches - could be like. Enhanced by the study of scientific works on information systems and education sciences, the following model offers an arrangement of the main influence factors as well as their relationships. It does not, of course, claim to be exhaustive but simply tries to lay the foundation for theoretical modeling, which can subsequently be improved by further research.

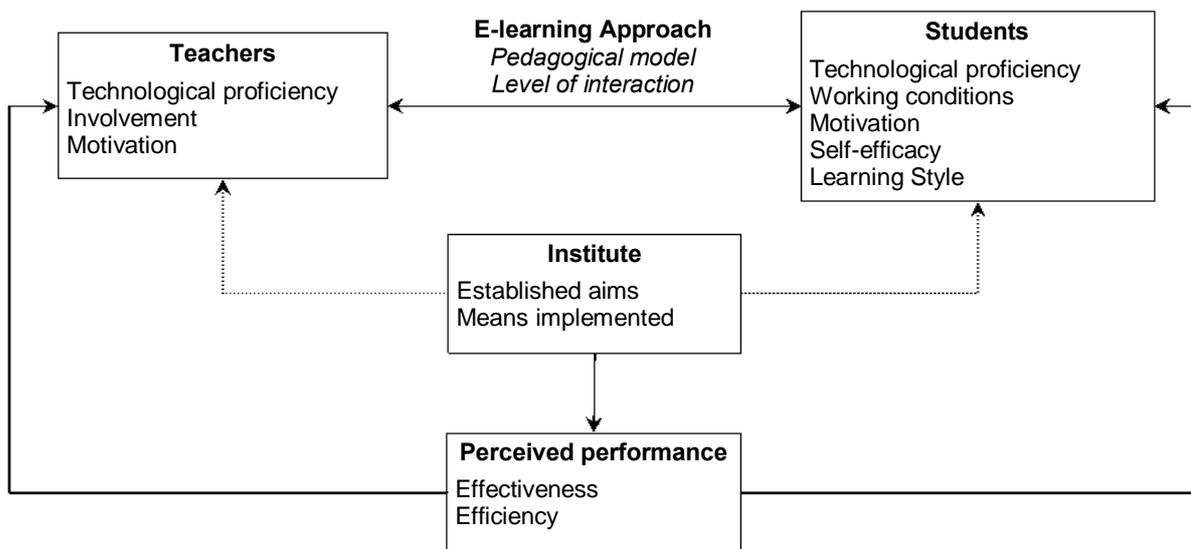


Figure 1: E-learning and performance flowchart

2.1. The e-learning approach

The first stage is of course to define the subject of our analysis and thus to determine the characteristics we retain from an *e-learning* approach. Indeed, rapid development of distance teaching, driven by spreading Internet technologies, has put a strain on neologisms

inducing multiple variations on the same theme. *E-learning*, *virtual class*, or even *digital campus* seem, at first sight, to be concepts whose main differences remain limited to the scope of the technological system installed and to the related pedagogical mode. The table below summarizes some of the definitions.

Table 1: E-learning definitions and similar concepts

	Description	Reference works
<i>E-learning</i>	Teaching method via a technological tool such as Internet, etc., allowing teachers to make all or part of their lessons available to a public of learners, and without any spatiotemporal restrictions. To a greater or lesser extent, <i>e-learning</i> systems often include means with which to: maintain an interaction with and follow-up on learners, personalize programs, assess or self-assess learner progress, etc.	Piccoli & al. (2001) Webster & Hackley(1997) Minnion & al. (2002) Tu & Corry (2002) Northrup (2002) Hirumi (2002)
<i>Virtual class</i>	Specific type of <i>e-learning</i> allowing the simulation of a teaching space representing the classroom via new technologies (interactive audio and/or audiovisual conversations, white board, assessment systems, group work, etc.) in addition to the benefits offered by the new technologies (asynchronous communications, file sharing, automated performance indicators, etc.).	Leidner & Jarvenpaa (1993) Hiltz (1995) Copolla & al. (2002) Bieber & al.(2002)
<i>Digital campus</i>	Generalization of the <i>e-learning</i> approach in a training institute's range of programs. In this case, the technological tools must also be able to deal with the associated administrative procedures (enrolment, school fees, course payment, etc.).	Work from the UO-MLR open university (Université Ouverte – Languedoc Roussillon). Case of Catalogne Open University

Because the technological tool implemented is supposed to become the main interface in the learning process, it clearly plays a central role in an *e-learning* analytical model. As an extension of the media richness theory,

some research projects reveal that the features this system makes available will render it more or less compatible with certain pedagogical models. (Webster & Hackley, 2001; Leidner & Jarvenpaa, 1995). In 2002,

research carried out by Minnion & al., drew up various models related to psychological and philosophical trends:

- *Rationalism*: it is considered that knowledge can be transferred to the learner who, for this to happen, has to use his/her power of reasoning.
- *Objectivism*: knowledge comes from experience. The learning process is seen as an accumulation of knowledge.
- *Constructivism*: knowledge is the fruit born from interaction and exchange between learners. The teacher plays the role of organizer / group leader.
- *Imagination*: knowledge grows according to the learner's ability to create and scrutinize.

Depending on the subject taught, the teacher's preferences or the *student learning style*, one model is likely to be more suitable than another one. However, as yet, we do not have a scientific corpus enabling us to predetermine the efficiency of each one in different context. But the real danger reside probably more in not adopting any pedagogical model at all. In any event, this is what many case studies have highlighted in unsuccessful experiments noted (Piccoli & al., 2002). Thus, strictly speaking, the content conveyed or the variety of matching technological features is not what guarantees the tool's effectiveness, but rather their appropriation in the context of one or several pedagogical models.

The features offered by new technologies (such as forums, chat rooms, team working tools, etc.) aim to afford learners a more active role in acquiring or developing their knowledge. Indeed, interaction is considered as one of the key factors in catching and holding the attention of a "virtual public" liberated from the classroom workplace (Webster & Hackley, 1997). However, the degree of interaction between students and teachers depends on how the two parties react.

In a conventional class lesson, the interactions affecting the students' attitude and performance take place spontaneously, in real time. Teachers interpret student behavior, answer questions, clarify concepts, give preference to discussion, structure the lesson according to the time frame, etc. This ability to initiate and facilitate such interactions, to encourage or steer *feedback*, is what, among other things, characterizes the teaching profession. In an *e-learning* approach, communications are mainly asynchronous and, through technology, are media-based. Opportunities to interact in real time are generally confined to planned sessions which can be incorporated into the pedagogical approach at certain precise moments. It is therefore particularly useful to classify the possible types of interactions when using the *e-learning* solution (Hirumi, 2002)¹.

The typology put forward by Moore (1989), widely used since then in education sciences, classifies interactions into three types, in relation to the sender and receiver: *learner-learner*, *learner-teacher* and *learner-content*. Let us look at each of these in more detail.

The *learner-content* interactions are defined as a process that consists of "*intellectually interacting with the content so as to bring about a change in understanding of the topic on the part of learners and to enhance their cognitive structures*" (Moore, 1989). Even when the students are alone, they have to commit themselves to this type of "internal" dialogue so as to encode and retain information (Berge, 2002). Content can only become knowledge for the student if this active cognition process occurs. In most cases, in a learning situation, content quickly becomes inert if there is a lack of immediate practical application ensuring this cognitive acquisition. It seems that making knowledge and skills available, just before an opportunity to use them, makes learning more effective (Gagné, Yekovich & Yekovich, 1993). This means it is vital to regularly alternate theoretical presentations with practical sessions in the courses offered (case studies, multiple-choice tests, etc.). This just-in-time requirement raises a certain number of questions concerning the simplification of the chosen cognitive approach and its long term effect on the student's skills.

The *learner-learner* interactions take place individually or in a group, with or without an instructor (Moore, 1989). As Northrup (2001) reminds us, *e-learning* is characterized by being able to learn anywhere and at any time via information technologies, but which may isolate the student. To overcome this feeling, group work or any other form of team work is often recommended. Forming social bonds and fulfilling part of the work as a team is not the only role of these learning activities. They also make it easier to reach the goal in terms of knowledge acquisition. The analysis of learning communities (Tu & Corry, 2002) shows us that besides individuals, the community learns even in a remote context. Vygotsky's research (1978) demonstrates just how important the social context is, and in particular interactions between learners, in the learning process. The environment should therefore stimulate exchanges between learners enabling beginners to take their shortcomings on board and to change their views through communication.

The aim of the *Learner-Teacher* interactions is to motivate and stimulate learners by allowing them to clarify concepts introduced in the content (Moore, 1989). One of the teacher's roles is therefore to interact with students, so as to help them overcome any difficulties that could not be resolved through contact with the subject matter or with other students. In addition, the teacher has to check the program is running smoothly. To do this, he/she ensures the program is being followed correctly by examining the path taken by each learner (time spent on the lesson, etc...) and lends guidance if a problem is identified. Tools such as individual forums or exercise correction can back up these interactions.

¹ Here we refer the reader to an education science literary review by Bannan-Ritland in 2002 regarding the concept of interaction in the e-learning field. We have only adopted the most accepted concepts.

Regular *feedback* often adds to student satisfaction with regard to this learning approach (Northrup, 2002). In this way, the teacher also strives to manage each student's progression.

A teacher, who consequently has to manage both teaching and learning, has to create an environment in which learners can be involved in projects, problem solving and other activities. The teacher is not a spectator but rather a co-discoverer who guides the learning process and encourages different types of interaction as well as reflection. Interactions are at the heart of learning, the success of which depends on the alignment between the aims, activities and feedback opportunities (Berge, 2002). In practice, the approaches aiming at improving these three types of interaction are those which are most likely to contribute to the success of e-learning². These are the three explicit axes of success for our model, with the primary objectives being all those outlined by the school, the activities and the feedback which mainly takes place between the students and the teachers.

2.2. The Teachers

Introducing e-learning can require the teacher to make major changes and even transform his/her way of teaching (Jean, 2001; Copolla & al. 2002; Godinet & Caron, 2003). Exercises are no longer tied to the singular space and time frame in which they were confined. Content often provided orally has to be pre-formalized (in writing, audio, video, etc.). A separation of roles between lesson designer, tutors and various experts, means the course design shifts from *handcrafting to mass-production* and from *individual to group*. "Stage plays" combining verbal and non-verbal communications are replaced by more impersonal contact (if we refer to the richness of media theory in any case).

The pedagogical style also requires a transformation. The teacher moves from the position of holder of knowledge or facilitator in student development to a role of regulator. In fact this migration, *from "Sage on the Stage" to the "Guide on the Side"* introduced by Copolla & al., 1997 does not eliminate the various roles that a teacher is expected to play. Following a series of around twenty interviews, the authors conclude, for example, that virtual teachers continue to exercise their cognitive, emotional and even domineering roles. To do so, they have to develop new behavior and to be able to use information technologies to communicate some of those signals (via forums, Email, etc.).

Therefore, if teachers wish to capitalize on the potential offered by *e-learning*, they need to be able to manage information technologies. Even if assistants can relieve them of technical tasks, they should be able to interact

directly with learners and follow-up on their work via the platform, to make practical information available on-line, etc.

Similarly for information technologies in an organization (Sproull & al., 1987), all these reasons mean that introducing *e-learning* in an establishment can induce a form of anxiety and influence the level of motivation. For all that, some empirical research reveals that the teacher's level of involvement in *e-learning* is a decisive factor in its success (Piccoli & al., 2001; Webster & Hackley, 1997). These conclusions confirm the theory of social influence related to the use of technologies (Fulk & al., 1990) according to which behavioral models developed by some are based on behavior observed by others. The case of *e-learning* appears to make this issue even more striking as teachers and students hold asymmetrical positions and the former are supposed to act as role models for the latter.

In addition to being sufficiently familiar with information technologies, *e-learning* also necessitates a positive attitude towards them. Training teachers how to use them may simply not be enough to develop this type of culture.

2.3. The Learners

Teacher characteristics, for all that, would not be enough to predict the motivation and the active behavior that the students will develop. Going beyond the vigor attached to *e-learning*, we should not neglect the feeling of frustration or isolation that distance learning can exert on individuals (Hara & Kling 2000). The more virtual an organization becomes, the more users tend to need face-to-face encounters (Handy, 1995). In addition to being motivated for the lesson, students also have to be motivated to learn via the *e-learning* mechanism in relation to their own command of information technologies and how to use them.

This form of teaching therefore also involves a cultural change for the learners. They are required to develop a more active behavior, to access knowledge in a more open information space, whereas up until then they had been used to receiving it in the confines of a classroom. Therefore, learners have to acquire a high degree of autonomy, which is more supposed than facilitated by the e-learning mechanism. The e-learning approach is supposed to provide learners with greater freedom, but at the same time they have to be able to envisage their own self-efficacy. Even if follow-up monitoring by teachers appears as a prerequisite to learning (Piccoli, 2001, p. 8) it seldom include the control of the way students organize their work and manage their "virtual timetable".

Scientific and professional literature has, at length, dealt with pedagogical models that should be adopted in an e-learning activity. However it has often be done to the detriment of learning styles. This is probably due to the fact that this student characteristic cannot be known

² Report for the Canadian government in 2002: "Methods and strategies for promoting on-line interaction for students undergoing distance learning."

beforehand and that there can potentially be as many different learning styles as there are learners. On the other hand, some Education Science research (Honey &

Munford, 1992) identifies 4 main styles which seem pertinent to us to take into consideration when setting up an e-learning system and its performance model.

Table 2: characterization of learning styles

	Characteristics	Preferences in terms of <i>e-learning</i>
<i>Thinkers</i>	They base their learning on data collection, observation, and listening to others. They analyze before drawing any conclusions. In group work, they mainly play a support and advisory role. Thinkers like to have sufficient time to observe and prepare.	Communication and group work tools Documentary sources and links to other references Self assessment tools
<i>Activists</i>	They appreciate new experiences, tending to apply concepts directly to them and subsequently inferring consequences and points of interest from them.	Group work communication tools Project based on content offering room for creativity rather than instructions to be followed. Short-term work Exercises, simulations, project fulfillment
<i>Theorists</i>	They use a rational approach in their cognitive process. Their approach is procedural and analytical leaving little room for an emotional intelligence.	Preference for individual work rather than problem solving. Clear pedagogical objectives and learning methodology.
<i>Pragmatists</i>	They need to see in advance what empirical use and implication a piece of knowledge or technique will have in their professional activity. As down-to-earth people they need to tie theory and practice together.	Aims and lesson plans clearly identified. Practical guides Feedback from experience gained in the field, from expert opinion.

Besides the pedagogical models advocating active participation by students, from now on we can consider that the e-learning lesson can also be given in any number of ways thus covering the full spectrum of mentioned learning styles. However, this is not sufficient to deduce that it would be a good idea to allow a student to enroll solely for pedagogical models supporting his/her own learning style or personal preferences. For example, it could be necessary to initiate and train a *theorist* student in working group methods. Consequently, it is better to think in terms of management of the different learning styles than, simply, in terms of their applicability.

2.4. The Institute

In scientific literature, the institute in which the e-learning activity is set up has a low profile or is totally eclipsed with respect to the proposed models.

However, the little research done emphasizes the organizational upheavals induced by e-learning as well as the subsequent and compulsory input from the institute (Jean, 2001). As previously highlighted, recorded field studies have mainly been proven experimentally between teachers and learners using a specific technological tool (*all things being equal otherwise*). In an era in which adopting information technologies represents a strategic challenge for schools, in our opinion, it is time to analyze the 'school-related' variables likely to play a role in the success of an on-line teaching activity.

Theories concerning the use of new technologies evoke the fact that their spread is a factor in their own acceptance. Spread is the process by which the technology is extended to other parts of the organization

(Goodman & Sproull, 1990). Opportunity is created for others to use this technology and to be aware that others use it. This spread is required in order to create a prescriptive general opinion of the new technology. Upheavals in the teaching activity brought about by *e-learning*, mean that its implementation depends on how determined a school is and resources available.

Adhering to a certain technology depends on individuals' determination, but also on how the management team "promotes" the idea (Salanick, 1977). And yet, there is a "values paradox" here (Sproull & Hofmeister, 1986): the more the technology is emphasized, the more severely it will be judged, if the pre-stated aims are not achieved. Adopting a new technology also depends on the symbolism associated with it (Prasad, 1993). This symbolism can be at the root of resistance or over zealous use. It also influences the system set up, thus becoming one of the primary reasons for adoption. For instance, as a symbol, modernity is a driving force in the spread of information technologies. This symbolism is the result of the combined effect of management and socio-cultural variables related to the field of application (higher education in our case). The way in which an *e-learning* project is introduced to the players involved will therefore also be an influential variable in terms of the perceived level of success related to the measured efficiency.

On-line learning brings about major changes in the teaching profession requiring just as much investment from the school. We should in particular mention:

- The incentive system for teachers: provisions regulating terms and conditions of employment for teachers is still out of phase with the rapid development of distance learning. For instance in the French university system, an hour's teaching is likened to a

service that has to be carried out, in due course, in a classroom and in front of students. This raises the problem of payment for e-learning type lessons and incentive measures taken to encourage teachers to put in an effort.

- Training for teachers and help in designing on-line lessons, multimedia resources, quizzes, student follow-up, etc.
- Forming an editorial committee entrusted with assessing teachers' work: meeting pre-determined quality criteria, abiding by copyright laws, etc.
- Media team responsible for transforming the resources developed by the teacher (web, flash, audio, video formats, etc.).
- Technical team responsible for putting the e-learning platform on-line and up-dating it.
- Etc.

Of course, all these factors determining the success of an e-learning tool, as well as those already presented in this article, are by no means an exhaustive list. The organizational complexity of educational establishments makes it difficult to predetermine the success or failure of any given project. This being said, we have the advantage of being able to analysis the success of a student in a learning situation using a relatively clear performance indicator, that of exam results.

2.5. Success of an E-Learning System

The main scientific research into experimenting with such systems (Bieber & *al.*, 2002; Minnion & *al.*, 2002; Coppola & *al.*, 2002; Piccoli & *al.*, 2001; Webster & Hackley, 1997; Hiltz, 1995; Alavi, 1995 & 1994) have analyzed the effects induced on learners and teachers in the following ways:

- learner-teacher interactions
- group exchange between learners themselves
- cognitive processes and pedagogical models
- cultural changes
- experience gained by the teacher and learner
- etc.

This experimental methodology-based research, often conducted on a test group of learners, has thus allowed the strengths and weaknesses of these tools to be brought to light - with regard to teaching and learning processes. It reveals that a multiplicity of influence factors confers a particularly subjective character to the idea learners and teachers may have of it. This can be explained through the differences in appraisal that each one makes of the: (1) pertinence of the set aims (purpose or acceptability concept), (2) relationship between the aims and the results achieved (effectiveness or usefulness concept) and (3) relationship between these results and the means (or effort) employed (efficiency or usability concept)³.

³ The *purpose*, *effectiveness* and *efficiency* definitions, frequently used in Economic Sciences, are reiterated here (see the Caby & *al.*, 1999 article to this effect).

If the arguments in favor of trying *e-learning* are, initially, with a view to improving the training process, others are interested in the underlying economic and strategic potential for an institute in taking up such a project. In this connection, another perspective is to consider *e-learning* as a way to rationalize the organization's operating costs (Minnion & *al.*, 2002). If we deem that the knowledge to be transferred to learners and the interactions associated with their understanding can be formalized via a computer-communication process, then the hypothesis of reducing the operating costs of a teaching activity is a legitimate one. All of the following should, at first sight, contribute to reducing the cost of teaching per student and generate savings likely to secure a return on the technological investment: an automated didactic system, self-assessment functions, teachers called on essentially for pedagogical follow-up, less classroom space needed due to face-to-face lessons, etc. Some research even associates this reasoning with an emphasis on development or even as a competitive edge acquisition (Webster & Hackley, 2001; Dufner & *al.*, 1999).

The e-learning market nevertheless entered troubled times two years ago. Caution needs to be taken therefore, in terms of the amount of ambition it is suitable to attach to research into the degree of success of such distance teaching tools. The fact that some "e-training" centers have recently gone bankrupt warns us that viability of an economics model based on information technologies should surely first transit by the appropriation and use of the latter.

So without trying to test the hypothetical impact of e-learning on the reduction in operating costs or the creation of a competitive edge, we believe it is nevertheless also necessary to take the idea of performance from an institute's standpoint into account. Is it still coherent, in fact, to consider that an organization's investments in information technologies (regardless of the organization) are pertinent only if they form part of a growth strategy (Powell & Dent-Micaleff, 1997)? On the one hand, the return on investments is not always assessable on objective grounds:

- certain costs cannot be measured accurately (for example: the hourly production cost of creating or transforming a lesson)
- the context, between when the decision is made to invest and when the system actually bears fruit, can radically change
- labor regulations as yet do not cater for distance learning (case of the French university system)
- etc.

On the other hand, the decision to invest may also be motivated by less analytical ambitions in financial terms. For several years now, training centers have been

We have also drawn a parallel with the Education Sciences research on the *acceptability*, *usefulness* and *usability* idea (see the Tricot & *al.*, 2003 article to this effect).

subjected to a certain pressure concerning experimenting with or adopting distance learning systems (Webster & Hackley, 1997; Alavi & *al.*, 1995). Implementing them can therefore aim at promoting a modern image or even avoiding being marginalized in relation to developments in the field. Measuring effectiveness and efficiency as perceived by the institute's management team may also bear fruit in terms of overall success.

However, if we take the stance that the primary purpose of e-learning is to offer learners a working platform allowing them to improve their results (Maki & *al.*, 2000; Schutte, 1997; Hiltz, 1995; Webster & Hackley, 1997), it is appropriate to analyze the efficiency of the said potential in a real situation and context that goes beyond a mere simulation or experiment. The empirical part of the research in progress, for the moment, will concentrate on analyzing the results of an entire year group of students (400 learners) having taken five on-line courses as part of their syllabus and throughout the academic year.

3. RESEARCH FIELD

For the moment, the data collected only allows us to make a comparative analysis of the results of the students having taken these on-line courses with the results of those who had previously taken the same courses by traditional methods. Tests of influence levels of the depicted characteristics and factors (student satisfaction level, teacher involvement, pedagogical model, etc.) will be the subject of subsequent publications.

3.1. The Institute

Montpellier Business School's *e-learning* project was given impetus in early 2001 with the decision to make it compulsory for students following the Business School program to do a year's study abroad. Consequently, at the start of the 2002 academic year, 400 students were dispatched to 130 different foreign partner universities. In addition to the lessons they attended in their host universities, these students had to follow some Montpellier Business School courses via the e-learning platform developed for this purpose. The challenge for the students was to obtain two diplomas in the same year. In any case, these courses had to be followed and validated for a student to be able to move on to the next year of studies. This is why group work was encouraged to complete assignments which required coaching by tutors as part of internal assessment.

For all that, e-learning was developed here in a bid to open the training program to the rest of the world. In this way, the primary aim was not as much to develop the pedagogical tool as to take full advantage of the distance learning opportunity. The following means were employed to achieve this goal:

- recruiting a researcher-teacher as person in charge of the e-learning project,
- enrolling the services of a computer engineering consultancy firm,
- training the teachers involved,
- implementing an editorial committee (made up of teachers representing the education and research departments) responsible for validating each teacher's work..

As an incentive and to give value to the teachers' work, a reduction in their teaching load was implemented:

Table 3: incentives for creating *e-learning* lessons in the case studied

<i>The year the e-learning lessons were created</i>	<i>Presenting and up-dating lessons in subsequent years</i>
<p>Reduction in the teaching load equivalent to the number of face-to-face teaching hours as per the course syllabus (disregarding the different year groups involved). <i>Example: reduction of 30 hours for a 30-hour e-learning course to be taught the following year.</i></p> <p>Reduction of 100 hours in non-face-to-face teaching time. This reduction concerns tasks of the following type: taking part in student selection panels, managing cross-referenced projects, etc.</p>	<p>The e-learning teaching time that students have to fulfill is counted in the same way as the time spent attending face-to-face teaching periods. <i>Example: if in a 30-hour course, a student year group has to spend 10 hours e-learning, these 10 hours will be counted in the annual teaching load that a university teacher has to provide.</i></p> <p>This time is, in particular, to be spent keeping the course up-to-date.</p> <p>A reduction in non-face-to-face teaching time (attributed as stipulated above) that the teacher should assign to giving the e-learning lesson (particularly for: answering the questions on the forum, posting practical information about the work expected, etc.). This reduction, on an annual basis, is equal to:</p> <ul style="list-style-type: none"> • 100 hours. • If fewer than 400 students, the reduction is calculated according to the following weighting: 0.25 hour x number of students enrolled for the on-line course. <p>In order to offer favorable teaching conditions (to organize certain modules, to up-date certain data, etc.), the teacher benefits from a teleworking scheme:</p> <ul style="list-style-type: none"> • One day a week (a set day each year) he/she is exempt from having to be at Montpellier Business School. • A high-speed Internet connection is made available at the teacher's home (DSL, cable or Numéris line). • A mobile phone with a subscription for one hour's communication per month designed to cover his/her professional calls.

This system of lightening the workload and related obligations has been formalized in the employment contract of all the teachers involved.

3.2. The E-Learning System

Emphasis has been put on an approach and a system affording the greatest possible flexibility to an innovative project whose specific needs had not been predetermined. Instead of buying an existing ready-made platform, the decision was taken to build a “homemade” one using open-source technologies. Developed by the group’s Webmaster, this platform, which is now in its fourth version, offers the same features as a professional product available on the market (content chunking, discussion forum, schedule for work to be handed in, glossary, quiz, etc.). This choice is justified by the fact that total control over the source code allows changes to be made at any time and specific needs met.

The e-learning platform includes, in particular, a system allowing the teacher to create lesson content on-line him/herself with the same ergonomics and the same functionalities as a conventional word-processor (copy, paste, text layout, inserting pictures or animated objects, etc.). As a result the teacher does not need any special IT skills.

3.3. The Learners

The 400 students on a year’s study abroad were hosted in a variety of universities in which they had computer rooms at their disposal allowing them to make full use of the e-learning platform functionalities. Incidentally, the only material needed was a computer (PC or Mac) connected to the Web with an Internet browser (and perhaps a printer if the student wanted to keep a hard copy of the lessons rather than viewing them in their electronic format). The questions or complaints made by the students (often by Email) were more oriented towards the conditions and difficulties of reconciling two courses at the same time, as opposed to problems with using Web technologies.

Table 4: main features of the *e-learning* tool implemented in the case studied

<i>Teacher environment</i>	<i>Student environment</i>
Zone for direct content design (put on-line according to an advanced WYSWYG ⁴ -type) allowing the teacher to create lesson content on-line him/herself.	Access to lesson content in the form of Web chunks and downloadable versions
System for creating a Quiz on-line with feed-back allowing the student to be directed in particular to parts of the lesson that need to be reworked.	Access to practical information (timetable, standard procedure instructions, etc.),
System giving the right to a student year group to access all or part of the lesson to be selected.	Access to “subject + correction” type exercises or self-assessment
Access, in his/her capacity as moderator, to his/her lesson forum: answering students’ questions on-line and managing messages	Access to a forum allowing the students to ask the teacher questions and share information and experiences among themselves. Practical tools: search engine, glossary, links, etc.

In terms of IT skills, these students could all be considered as having a satisfactory level for using the e-learning tool set up:

- in their first year of study on the Montpellier campus, they all had about 30 hours of compulsory computer science lessons,
- by the way, the platform requires no other skill than knowing how to use a Web browser.

3.4. The Teachers

The 5 teachers involved in the school’s project covered the following subject areas: Finance, Process Management, Auditing, Information Systems, and E-business. Each teacher was responsible for drafting the content of the lessons to be converted to *e-learning* material. To do so, they had to adhere to the following pedagogical model:

- Cut course content up into sessions equivalent to 2-hour-long face-to-face lessons
- Divide each of the sessions into “chapters” or units of learning of no more than 3 screen pages. The basis writing reference to be adhered to, in terms of volume, is 20 A4 pages per session.
- Set the students a case study type assignment (in groups of 5) covering all the sessions developed and as part of their internal assessment.
- Compile the glossary, reference bibliography as well as the links to other recommended websites.

The work submitted by each teacher was assessed by an “editorial board” before being authorized to go on-line in September 2002.

⁴ “What You See Is What You Get”

The teachers were held by institutional accountability to ensure student follow-up. Given the number of students to monitor (400) scattered worldwide in different time zones, asynchronous communication proved to be the most appropriate vehicle for their interaction. It was also important for the school that the chosen system be as user-friendly as possible and not require any specific software to be installed on the 120 host university computers. This is why, the use of electronic forums relayed to Email (for any confidential communication) was given preference.

3.5. Analysis of the Results

The diagrams below compare the final examination results⁵ of two year groups:

- (1) the 2002-2003 promotion, having followed all 5 courses on line
- (2) the previous year's promotion having attended class and followed these courses in the traditional way on campus at Montpellier Business School.

These 5 examinations were held on the same day, at Montpellier Business School Co., upon the students' return from abroad. They were in the form of multiple-choice tests containing the same number of questions / answers for each course and drawn from a database having been used by the teachers for several years. The level of difficulty was therefore similar to that of the previous year. Marking was not done by the teachers, but by the group's management team, with a scanner and software enabling the papers to be processed automatically.

As for the diagrams, they reveal a drop in the level in 4 out of 5 courses. Future research will concentrate precisely on collecting and analyzing data (both qualitative and quantitative) allowing these results to be explained. However, there is already an interesting fact that is worth our attention...

In terms of skills, none of the teachers had any special prior expertise in *e-learning*. And in fact the auditing teacher (the only subject to show an improvement in results compared to the previous year) showed a appreciably inferior command of information technologies than her 4 colleagues. The latter, for example, were used to: creating PowerPoint presentations, having their students search through reference sites on the Web or even giving exercises using specific software (Excel, MsProjects, Software Engineering Environment, etc.). The teacher in question declared not having used information technologies in her lessons up until then, and feeling more at home in the oral teaching mode (traditional lesson) than one that requires everything to be formalized in writing (*e-learning* lesson). In other words, the only subject

recording a higher level of student success is linked to the teacher with the least proficiency in new information and communication technologies.

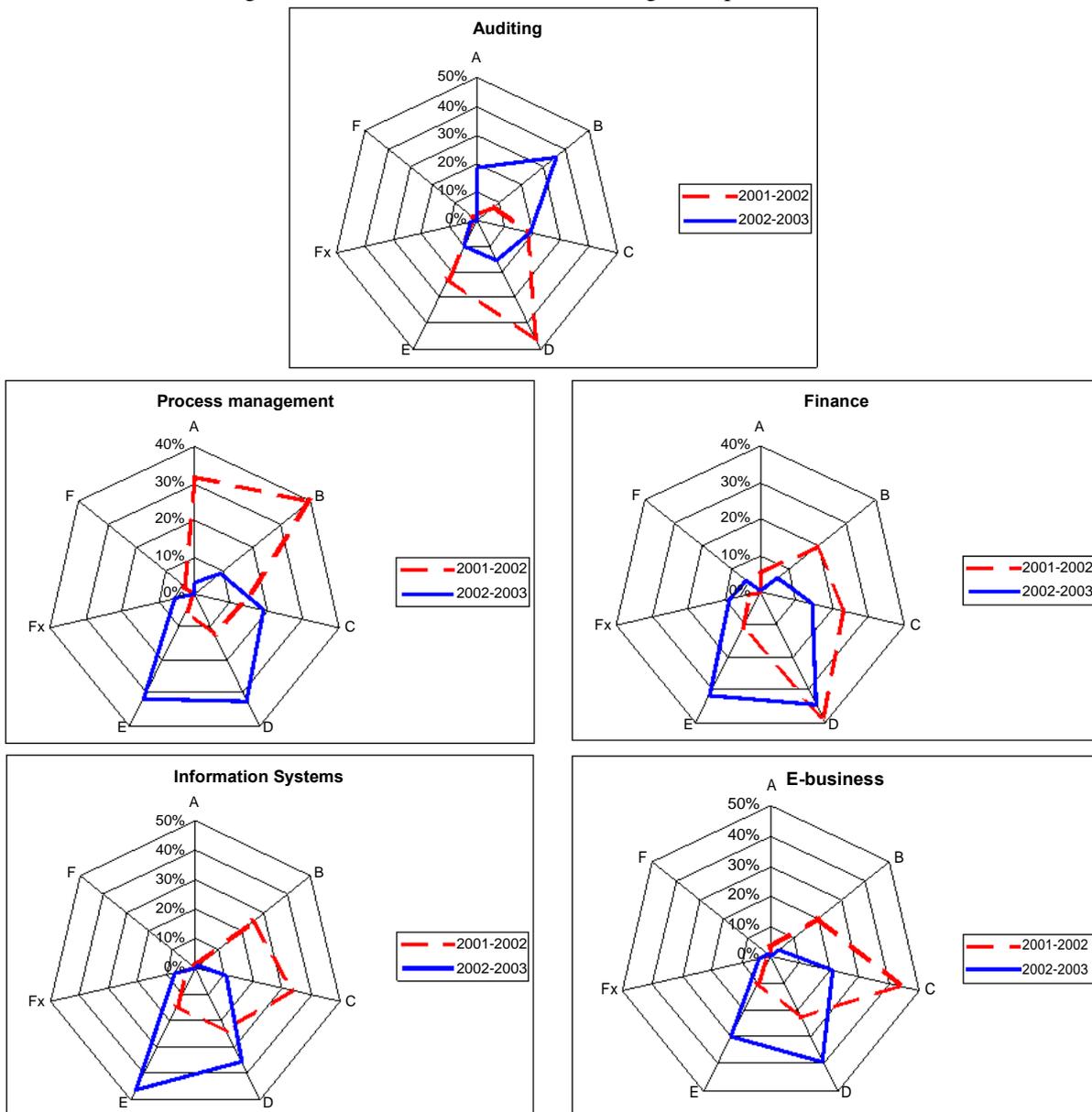
Consequently, this case, while unique, is it not enough to show that even if being conversant with information technologies is necessary for the "virtual teacher", it is not a determining factor in the success of an e-learning course? Is this not a perfect illustration of the need to dismiss professional reasoning as scientific in a techno-centered approach to on-line education?

On the one hand, given the diagrams, we could assert that the primary aim of such a tool has not been achieved. For all that, is this observation not worth putting into perspective so as not to adopt a simplistic approach to the legitimacy of this type of project? Indeed, this is the very first *in vivo* experiment for an institute, its teachers and students. These results can be interpreted as the formalization of the appropriation issues and cultural changes inferred by *e-learning* such as those we depicted in the literary analysis. For this reason, we believe that a methodology applied to the subject would be beneficial on a temporally longer system of reference, allowing the influence of this type of variable to be diluted.

For all that, over and above this decline in performance, the management team of Montpellier Business School considers the *e-learning* project as a success in as much as its implementation enabled the group to achieve the pre-determined globalization objective and to enroll an entire year group of students in a year's study abroad culminating in a double diploma. For this reason, the success of the tool must not be limited to examination results alone, but should include other indirect indicators. In this field example we see that the success of *e-learning* is closely linked to the success of the upstream project requiring its implementation.

⁵ The results here are detailed according to the marking schedule applied in the target universities: A = Excellent; B = Very Good; C = Good; D = Fair; E = Unsatisfactory; Fx = Fail (with the right to re-sit); F = Fail (Student excluded)

Figure 2: Traditional class *versus* e-learning – comparison of exam results



4. CONCLUSION

Cross-referencing two fields, information systems and education sciences, provides the opportunity to offer an explanatory model of the effectiveness of an e-learning tool. This model incorporates the teachers' and students' characteristics, but also the institutes' aims and means as well as the fundamental relationships between these three categories of player. The research carried out at the Montpellier business school initially only allows the results students obtained in traditional training to be compared with those gained in on-line training, which represents only the beginning of research intended to delve more deeply into the proposed model. However, these initial results support our reasoning: a view overly centered around technological determinism should logically be dismissed, in the same way as limiting the measure of success to the students results alone. The

aims of the school studied in this article show that the legitimacy of an *e-learning* project can lie more in the satisfaction of being able to meet new strategic challenges through its development than in simply improving an existing teaching tool.

5. REFERENCES

[1] ALAVI M (1994), "Computer-Mediated Collaborative Learning: An Empirical Evaluation", *MIS Quarterly*, Vol. 18, N°2, 1994, pp. 159-174.
 [2] ALAVI M., WHEELER B. C., VALACICH J. S. (1995), "Using IT to Reengineer Business Education: An Exploratory Investigation to Collaborative Telelearning", *MIS Quarterly*, Vol. 19, pp. 293-312
 [3] BANNAN-RITLAND B. (2002), "Computer-Mediated Communication, Elearning, and Interactivity, a

- Review of the Research”, *The Quarterly Review of Distance Education*, Vol.3, N°2, pp. 161-179
- [4] BERGE Z. L. (2002), “Active, Interactive, and Reflective Elearning”, *The Quarterly Review of Distance Education*, Vol.3, N°2, pp. 181-190
- [5] BIEBER M., ENGELBART D., FURUTA R., HILTZ S. R., NOLL J., PREECE J., STOHR E. A., TUROFF M. & VAN DE WALLE B (2002), “Toward Virtual Community Knowledge Evolution”, *Journal of Management Information Systems*, Spring, Vol. 18, N° 4, pp. 11-35.
- [6] CABY L., GREENAN N., GUIESSAZ A. & RALLET A. (1999), « Informatisation, organisation et performances des entreprises : quelques propositions pour une modélisation », dans D. Foray & J. Mairesse, *Innovations et Performances*, Editions de l'EHESS
- [7] COPPOLA N. W., HILTZ S. R. & ROTTER N. G. (2002), “Becoming a Virtual Professor: Pedagogical Roles and Asynchronous Learning Networks”, *Journal of Management Information Systems*, Spring, Vol. 18, N° 4, pp. 169-189.
- [8] CRADLER J. (1997), “Summary of Current Research and Evaluation of Findings on Technology in Education”, *Working Paper*, Educational Support Systems, San Mateo
- [9] DUFNER, D., KWON, O., & HADIDI, R. (1999), “Web-CCAT: A Collaborative Learning Environment for Geographically Distributed Information Technology Students and Working Professionals”, University of Illinois at Springfield. 1999.
- [10] FULK J., SCHMITZ J. & STEINFELD C. W. (1990), “A Social Influence Model of Technology Use”, in Fulk J & Steinfield, *Organizations and Communication Technology*, Sage, pp. 117-141
- [11] GAGNE, E., YEKOVICH C. W. & YEKOVICH F. (1993), *The Cognitive Psychology of School Learning* (2nd ed.), New York, HarperCollins
- [12] GODINET H. & CARON C. (2003), "L'accompagnement dans le Campus Numérique FORSE : modalités et outils". *Actes de la conférence EIAH 2003*, Strasbourg, 15,16 et 17 avril.
- [13] HANDY C. (1995), “Trust and the virtual organization”, *Harvard Business Review*, May – June
- [14] HARA N. & KLING R. (2000), “Students Distress with a Web-Based Distance Education Course: An Ethnographic Study of Participants Experiences”, *Information, Communication and Society*, Vol. 3, N°4, pp. 557-579
- [15] HILTZ S.R. (1995), “Teaching in a Virtual Classroom”, *International Journal of Educational Telecommunications*, Vol. 1, No 2, pp.185-198.
- [16] HIRUMI R. (2002), “A Framework for Analysing, Designing, and Sequencing Planned Elearning Interactions”, *The Quarterly Review of Distance Education*, Vol.3, N°2, pp. 141-160
- [17] HONEY P. & MUNFORD A. (1992), *Manual of Learning Styles*, Paperback
- [18] JEAN T. (2001), “L’EAD : Facteur Déterminant de Bouleversement Pédagogique et d’Evolution Organisationnelle et Stratégique”, Colloque Agora / TICE, Grenoble
- [19] LEIDNER D. E. & JARVENPAA S. L. (1993), “The Information Age Confronts Education: Case Studies on Electronic Classrooms”, *Information Systems Research*, Vol. 4, N°1, pp. 24-54
- [20] MINNION M., AMAMI M. & BRIMBERG J. (2002), « Information Technology-based Learning. The Royal Military College of Canada Experience », *Actes de la conférence de l’Association Information et Management*
- [21] MOORE M. G. (1989), “Three Types of Interaction”, *The American Journal of Distance Education*, Vol.3, N°2, pp. 1-6
- [22] NORTHRUP P. (2001), “A Framework for Designing Interactivity into Web-Based Instruction”, *Educational Technology*, Vol 41, N°2, pp. 31-39
- [23] NORTHRUP P. (2002), “Online Learner’s Preferences for Interaction”, *The Quarterly Review of Distance Education*, Vol.3, N°2, pp. 219-226
- [24] PICCOLI G., AHMAD R. & IVES B. (2001), “Web-based virtual learning environment”, *MIS Quarterly*, Vol. 25, N° 4, pp. 401-426, December
- [25] POWELL T. C. & DENT-MICALEFF A. (1997), “Information Technology as Competitive Advantage: The Role of Human, Business and Technology Resources,” *Strategic Management Journal*, 19(5), 375-405.
- [26] PRASAD P. (1993), “Symbolic Process in the Implementation of Technological Change: a Symbolic Interactionist Study of Work Computerization”, *Academy of Management Journal*, Vol.36, N°6
- [27] SALANCIK, G. R., “Commitment and the Control of Organizational Behavior and Belief”, in B. M. STAW and G. R. SALANCIK (eds), *New Directions in Organizational Behavior*, Chicago: St. Clair Press, 1977
- [28] SPROULL, L. S. & HOFMEISTER, K. R., “Thinking About Implementation”, *Journal of Management*, 1986, 12, pp. 43-60
- [29] TU C-H. & CORRY M. (2002), “Elearning Communities”, *The Quarterly Review of Distance Education*, Vol.3, N°2, pp. 207-218
- [30] TRICOT A., PLEGAT-SOUTJIS F., CAMPS J.-F. & al. (2003). "Utilité, utilisabilité, acceptabilité : interpréter les relations entre trois dimensions de l'évaluation des EIAH". *Actes de la conférence EIAH 2003*, Strasbourg
- [31] VYGOTSKY L. (1978), *Mind in Society*, Cambridge, Harvard University Press
- [32] WEBSTER J. & HACKLEY P. (1997), “Teaching Effectiveness in Technology-Mediated Distance Learning”, *Academy of Management Journal*, Vol. 40, N°6, pp. 1282-130.

Electronic Disclosure and Financial Knowledge Management

Jerome Yen, Percy Yuen, Belinna Bai

Department of Systems Engineering and Engineering Management,
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China
{jyen, ltyeun, lbai}@se.cuhk.edu.hk, Tel: (852) 2609-8237, Fax: (852) 2603-5505

ABSTRACT

In this paper we report the benefits of using eXtended Markup Language (XML) to support financial knowledge management, which include indexing, organizing, association generation, cross-referencing, and retrieval of financial information to support the generation of knowledge. The current searching engines cannot provide sufficient performance, such as, recall, precision, extensibility, etc, to support users of financial information. XML is able to partially solve such problem by providing tags to create structures. XML provides a vendor-neutral approach to structure and organize contents. XML authors are allowed to create arbitrary tags to describe the format or structure of data, rather than restricted to a specific number of tags given in the specification of HTML. A prototype of XML-based ELeCtronic Financial Filing System (ELFFS-XML) has been developed to illustrate how to apply XML to model and add value to traditional HTML-based financial information by cross-linking related information from different data sources, which is an important step in moving from traditional information management to knowledge management. We compared the functionality of XML-based ELFFS with the original HTML-based ELFFS and SEDAR, an electronic filing system used in Canada, and recommended some directions for future development of similar electronic filing systems.

Keywords: Financial Knowledge Management, Electronic Financial Filing Systems, XML

1. INTRODUCTION

Corporate filings, news articles, trading data, information about the listed companies, etc, are some most frequently requested data by investors. Majority of financial data or information are disseminated through Internet hoping that they will be widely available. However, searching, retrieving, and utilizing financial data have become a major challenge to both users and researchers. Such challenge includes issues, such as, efficiency, precision, recall, relevance, and value creation. According to the 10th WWW User Survey, only around 25 percent of people used the WWW to access financial information on daily basis. The same survey also pointed out that more than 45 percent of the respondents felt that they were not able to find the information they needed. To solve these problems, content creation, representation, and management can be few areas to attack. It is possible to automate or semi-automate the process of content creation, collection, indexing, organizing, storage, and dissemination.

In this paper, we dissect the problems mentioned above into the life cycle of financial information and proposed a framework to use eXtended Markup Language (XML) to rectify these shortcomings. It is also hoped that it will help users to better manage financial *data* by systematically convert, synthesize, and create financial *knowledge* from *data and information*. It is a very important element in knowledge management [9]. To support this research, a prototype of XML-based ELeCtronic Financial Filing System (ELFFS-XML) has been developed to illustrate

how to apply XML to model traditional HTML-based financial information and providing value added services. We compared the functionality of XML-based ELFFS with the original HTML-based ELFFS and SEDAR and recommend some directions for future development of similar electronic filing systems.

For the rest of the paper, Section 2 provides a short discussion about related research works of financial filing systems as well as the issues and problems of organizing financial information on the Internet. Section 3 discusses the Financial Knowledge Management Framework. The implementation and major technologies of ELFFS-XML are discussed in Section 4. Experiment and user evaluation are discussed in Section 5. This paper is concluded with a discussion about future research.

2. FINANCIAL INFORMATION MANAGEMENT AND ITS ISSUES

Transparency and efficiency are two extremely important issues to the users of financial information [8]. Transparency can be defined as the visibility of a financial market and efficiency refers to the time and efforts that needed by users to search and retrieve the needed information. In order to make high quality decisions and narrow the gaps between private investors and corporate investors, both issues need to be solved.

Financial information/data were formerly disseminated through traditional media, such as five nationwide Securities and Exchange Commission (SEC) reading

rooms in the US. To obtain the needed information, investors had to travel to these locations, search, and then make photocopies, which was extremely inefficient. Private companies, such as Mead Data Central, also provided financial information via dedicated connections with extremely high charge that not every investor could afford.

Since the advent of the Internet and hypertext, most financial information were digitised and disseminated through Internet. In the following, we will discuss the existing Electronic Filing Systems and current financial information management issues.

2.1 Existing Electronic Filing Systems

As Internet appeared, a systematic solution – the electronic financial filing system was emerged to aid parties, such as, filers, regulators, and investors to collect, process, store, and organize financial information. Filers usually are the companies that need to file financial documents to the regulatory bodies, such as Securities Exchange Commission (SEC). They are required by law to submit information for disclosure to the public. The regulators are responsible for collecting the submissions, processing them, guaranteeing the information received are correct, valid, and sufficient, and making them available to the public.

Different countries have different electronic financial filing systems. SEDAR [3] is a Canadian project that focuses on building financial information repository to support investors. It collects financial reports or documents from Canadian listed companies, which can be in different formats, and creates a web accessible archive for users. EDGAR [2] (Electronic Data Gathering, Analysis and Retrieval) is the US efforts in supporting collecting, indexing, organizing, and disseminating financial filings through Internet. It provides facilities for filing parties to format their reports in SGML. EElectronic Financial Filing System (ELFFS), was an initiative at the University of Hong Kong endorsed by the Securities and Futures Commission (SFC) of Hong Kong Government to support digitising, collecting, archiving, and disseminating financial information using traditional HTML [8].

2.2 Current Management of Financial Information on the Internet

We have carefully analyzed and studied the strengths and weaknesses of each filing system, and found there are still lot of rooms for all the filing systems to improve. An ideal electronic filing system should allow collecting documents from submitters via Internet to increase the speed of transmission as well as lower the transmission costs. It should support processing and validating the submissions, store and create archives, organize, index, and extract the useful clues, trends,

and relationships buried in the data. This is a crucial step, which can be defined as knowledge generation.

From the day when the World Wide Web was born, a HyperText Markup Language (HTML) was needed to "mark up" contents in order to create the format that can be presented on the web. HTML, deals with two different views of the same piece of information: *what* the content is and *how* the content is presented. This is what *metadata* [4] is about. HTML documents are actually a mixture of both content and presentation instructions, which makes it very difficult to process by both human beings and machines. As some of the syntaxes used in HTML are not too strict, data extraction from a HTML document may not be even possible.

Over the past few years, Internet has become a major information source for investors and it provides information that covers almost all the major financial markets in the world. People started to create value from the huge amount of raw information, but processing it to create *knowledge* is difficult. Raw data exist in many forms, and when it comes to unstructured documents like HTML and all those generated by presentation-based software, content extraction becomes even more difficult. Quick adoption of Internet also accelerated the explosion of digital contents and information overload has become a major challenge. Data redundancy is another problem, e.g. one copy in the raw information and another copy in the presentable page. Keeping two sets of data not only cost more storage, but also cost more manage efforts.

2.3. A Survey about the User Issues

We conducted a survey to study the situation on the usage of financial information. 42 people are interviewed for the major problems they encountered when accessing financial information over Internet. 30 of them were investors, 10 were financial software developers and the remaining were information system administrators. Based on a scale of 1 to 10 (from least serious to the most serious), we found that the financial information on the Internet faced the following issues, the number in the bracket indicates the average score of severity:

The

1. **Too time consuming to gather** (8.13)
2. **Volatile to catch** (7.64)
3. **Distributed and fragmented in nature** (7.69)
4. **Bulky to store** (7.68)
5. **Too fast to collect** (7.54)
6. **Easily perishable** (7.49)
7. **Not unique** (7.34)
8. **Varying in content representation/format** (7.29)
9. **Difficult to be processed by human** (7.16)

3. FINANCIAL KNOWLEDGE MANAGEMENT

According to Thomas Davenport [7], knowledge management is a systematic attempt to create, gather, distribute and use knowledge. Matt Richtel [15] defines knowledge as information that is accessible and can be directly used to solve a problem while it is just raw and impenetrable. In simpler words, knowledge management is the methodology to add value to raw data and convert them into useful knowledge.

3.1 Financial Knowledge Management Framework

From the point of view of an information producer, we break up the Financial Knowledge Management into seven stages: *information creation, information collection, information storage, knowledge generation, knowledge dissemination, knowledge presentation and feedback*. It was derived from the knowledge management processes that discussed in previous section and was adapted specifically to suit the needs for Financial Disclosure and the development of the prototype (Figure 1). We have briefly categorized the information consumers into three groups, namely, individuals, financial regulators and value adders. Due to the scope of the paper, we will only discuss the needs of individual investors.

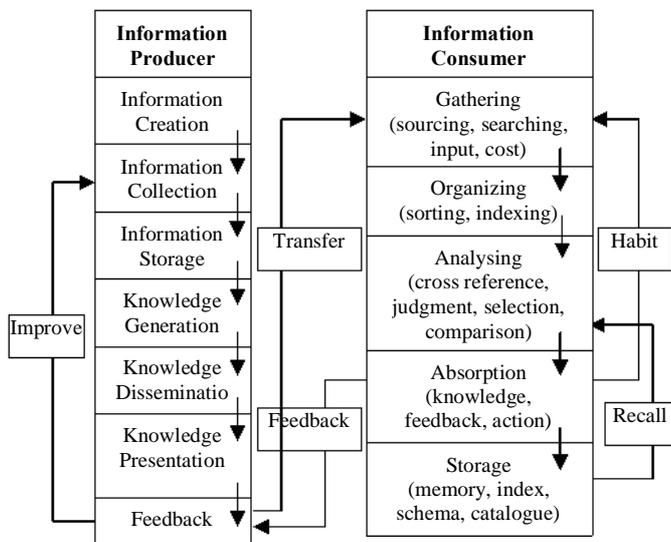


Figure 1: Financial Knowledge Management Framework

3.2. Design of the Knowledge-based ELFFS-XML

Based on the Financial Knowledge Framework, we designed the general architecture of ELFFS-XML (Figure 2). In the following, we will discuss in detailed tasks in the each layer of the architecture.

3.2.1. Information Creation

In ELFFS-XML, we aimed to provide facilities for content creators to create structured contents using XML. The first step was the conversion of the Java-

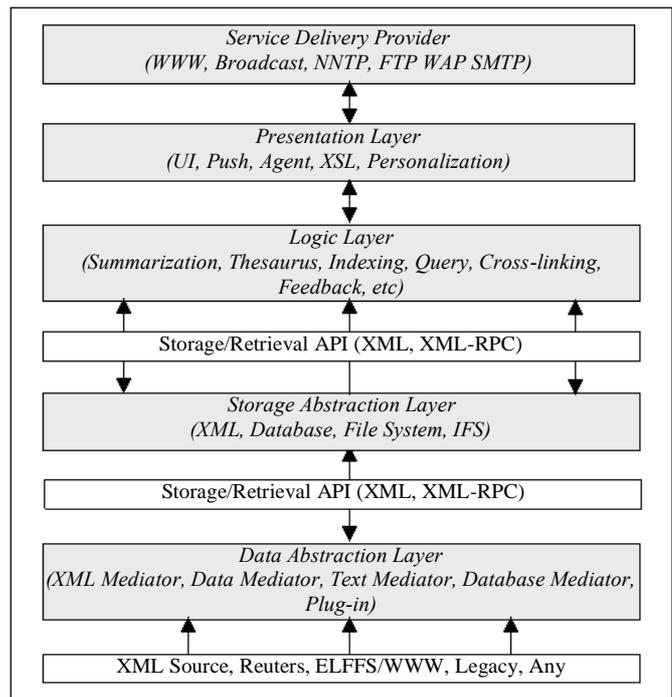


Figure 2: Architecture of ELFFS-XML

based cross platform Formatting Wizard for financial report files. Users can input the data using a browser without knowing the details of XML.

3.2.2. Information Collection/Storage

There are many financial information sources available. We have developed a set of retrieval agents that can connect to a variety of data sources including ELFFS financial report archive, Reuters, WWW, financial knowledge bases and other legacy sources. Each of them is working independently and the data they collect is converted into XML formatted files through a standard Application Programming Interfaces (API).

The XML files are placed in a traditional file system and an archive is formed gradually. They are recombined in real time using the "SYSTEM" entity feature of XML. Although it may increase access time, it actually increases the flexibility of handling these data. Until a general XML mass relational storage mechanism has been developed, we are staying with the primitive file system approach.

3.2.3. Knowledge Generation

We have a multiplexing engine that chooses from different kinds of data sources and invoke the predefined logic for each source, such as data summarization and cross-linking. An XSL style sheet is applied to an XML document that contains the annual report repository. Because XSL can be used to perform many operations on XML tags, it is equivalent to a logic layer. We used a Java-Servlet Cocoon to perform this XSL transformation on the server side.

3.2.4. Knowledge Dissemination/ Presentation

The Extensible Style Language is used again here. Using various tools, the XSL processor can output different well-presented formats such as the Portable Document Format (PDF), HTML, WAP pages or even XML.

4. IMPLEMENTATION AND MAJOR TECHNOLOGIES OF ELFFS-XML

In this section, we will discuss the implementation and major technologies of ELFFS-XML.

4.1 Document Format

The financial document is XML formatted. It starts with the <filing> tags. There are 2 major "sections" inside, <data> and <body>. Inside <data>, there are a number of vital data about the document, such as document identifier, document title, date, and a couple of keywords. Inside <body>, there are <sidebar> and <maintext> tags and inside <maintext>, there are <paragraph> tags, where content of the report is mainly placed. Some key performance data about the report have been tagged, such as profit last year (<perf-profit-lastyr>), profit this year (<perf-profit-thisyr>). After this XML page is fed into an XML processor, a XML document tree is generated. Document contents and data can be easily extracted.

4.2 Generation of Tags and Schemas

In reality, almost all the filers do not markup their file using XML. Even if they do, they are using different DTDs. This imposes a new XML standard called eXtensible Business Reporting Language (XBRL), which attempt to provide a framework defined by a standards group for XML business report. In this paper, we attempted a different approach. We tried to use a combination of three methods to generate the tags, namely, statistical and information retrieval approach, accounting and auditing practice approach and investor feedback approach.

4.2.1 Statistical Information Retrieval

A total of 5000 news article related to South East Asia available from Reuters, 200 governmental financial articles and 20 annual reports from different listed companies were pooled together to create the archive to support the identification of tags. We used several techniques to support tag generation. Two approaches are involved: the first one is calculating term frequency and document frequency of terms and phrases in all the documents; the second is using the concept space approach [11] as well as a couple of online financial and business dictionary to create our ELFFS thesaurus. Then we have performed object filtering and automatic indexing for each of our collected articles mentioned

above. A list of concepts was generated after we performed co-occurrence analysis. Term frequency and document frequency calculation were used to find out the combined weights of terms and phrases in the concept space. All the subjects in the automatically generated thesaurus were then used to help identify rotationally duplicated terms.

4.2.2 Accounting and Auditing Practise

To improve the reality and practicability of the meta data, we examined the code of practices of several accounting firms and auditing firms. An additional list of tags was generated after this process.

4.2.3 Investors' Feedback

Investors are usually the final recipient and consumer of the financial information. We have interviewed 20 investors about what information they want to see from different types of financial information.

5. EVALUATION OF ELFFS-XML

ELFFS-XML is evaluated by different criteria: system efficiency, tag generation performance and user satisfaction. Before the evaluation, we first take a look at how ELFFS-XML compares to other electronic financial filing systems.

5.1 Comparison with Other Financial Information Disclosure Systems

ELFFS-XML is compared with different such systems using the Financial Document Disclosure Model. The ELFFS and the SEDAR system were used as references (Table 1). It is not difficult to see that ELFFS-XML, in many ways, has trivial advantages.

5.2 Systems Efficiency

We have conducted a simple experiment revealing the size of an XML document can cost significantly more resources than a binary encoded document (store, network transfer, read, process.) We definitely need some ways to reduce the impact of this problem, but these are outside the scope of this paper.

5.3 XML Tag Generation Performance Evaluation

We conducted the experiment with the thesaurus agent (computer), a group of three experts in the field of accounting, finance and investment, and a group of five individual investors with different backgrounds. The thesaurus agent is running on a SUN-Enterprise 6500 server with 8 ULTRAsparc-II 400MHz CPUs and 3 GB of memory. Each of the experts and investors were given an accounting and a financial investment dictionary, twenty annual reports of Hong Kong, a list of twenty finance related web sites with a machine to

access those sites, a terminal that can access and view real time financial data from Reuters and 5000 pieces of archived Reuters news articles related to South East Asia market. After reading and examining these materials, the three groups of participants were asked to produce a list of tags they think useful. They are given a total of ten hours to come up with the list with

no limit on items. Both human and computer participants are only allowed to view web pages that are directed linked from the list of specified web sites to increase the fairness of the experiment. After the list was produced, the naming of the tags was adjusted so that they are consistent with each other in terms of semantic meanings.

	SEDAR	ELFFS	ELFFS-XML
Information Creation	extremely varied output format	formatting wizard, Java based, cross platform	formatting wizard, Java based, cross platform, keywords tagged, other XML aware editors supported
Information Collection	modem terminal upload	WWW secured via SSL	WWW secured via SSL
Information Storage	file systems based (usually 1 file), database	file systems based (usually multiple files), database	file systems based (usually multiple files), XML archive, optional legacy database support
Knowledge Generation	document indexes	document indexes, full text indexes	document indexes, full text indexes, cross referencing with online news articles, keyword search, multiple documents comparison and querying
Knowledge Dissemination	WWW	WWW, SMTP (email)	WWW, WAP, SMTP (email)
Knowledge Presentation	Word documents, proprietary formats, need special proprietary viewer	HTML	XML/XSLT, server side transformation to HTML, browser independent presentation, document map, PDF for printing

Table 1: Comparison of Different Financial Information Disclosure Systems

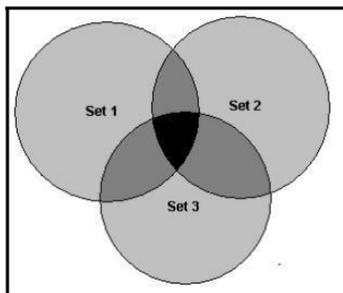


Figure 6: Different Sets of ELFFS-XML Tags

To quantify the performance of the different approaches, we view the selected tags from each of the three groups as separate sets as in Figure 6. Visually, the more dark area a set has with respect to its own size, the better is its performance. Mathematically, we use precision and recall as our evaluation metric. We slightly modified this concept and revised equations are:

$$P = \frac{\text{number of good tags produced}}{\text{total number of tags produced}}$$

$$R = \frac{\text{number of good tags produced}}{\text{total number of good tags}}$$

Visually, good tags are those tags inside the dark grey and black area. We used the F-measure to combine both ratings. The F-measure is defined as:

$$F = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$

$\beta = 1$ was chosen because we believe that precision and recall are equally important.

	# Correct	# Produced	Total Correct	P	R	F(1)
Investors	128	385	257	0.33	0.33	0.17
Experts	157	350	257	0.44	0.61	0.26
Agent	185	480	257	0.38	0.72	0.25

Table 2: Experiment Results for Different Tag Generation Approaches

6. CONCLUSION AND FUTURE RESEARCH

In this paper, we proposed a new approach to support indexing, organizing, and retrieval of financial information on the Internet using the eXtensible Markup Language (XML). The new approach includes models such as the document disclosure model, which was proposed to understand the underlying problems of information producers and consumers.

The strength of this new approach is that important data can be extracted relatively easily to facilitate easier creation of knowledge than the traditional HTML approach. ELFFS-XML can help to archive higher information transparency through the provision of filing standards and facilities to organize these information. End-users can gain benefits because they can obtain knowledge more easily. At the same time, it also provides multiple efficient delivery channels for end-users to receive the needed information at the lowest cost of time and money. It significantly bridges the gap between large corporate investors and small individuals, who are less resourceful.

We also found out that with the help of financial experts, computer agent, which automatically generate

metadata or tags from a multiple documents collected from different sources, can provide very good performance in terms of precision and recall. Despite how ideal it sounds, there are still works need to be done. Currently, the DTD defined for ELFFS-XML is in an experimental stage, which lacks sophistication and details. Many financial performance tags are still missing. We are looking forward to expand it into a full-fledged XML vocabulary and let government agencies evaluate its usefulness.

Scalability is another issue. Searching through a large text-base XML repository on a file-system-based storage can be very I/O intensive, which may become the bottleneck to large systems. Moreover, there are still no standard ways to query XML archive. All we have now is a XML parser and a set of procedures about how to use the parser. We are currently implement dbXML and XML-QL, which are SQL-like languages for XML, to improve performance and provide more features.

REFERECES

- [1]. Steve Lawrence and Lee Giles, "Accessibility and Distribution of Information on the Web", *Nature*, Vol 400, pp107-109, 1999.
- [2]. Ajit Kambil, "The Direct EDGAR Access System: Final NSF Progress Report", <http://www.stern.nyu.edu/~akambil/publications/nsfrep.pdf>
- [3]. "System for Electronic Document Analysis and Retrieval (SEDAR)", <http://www.sedar.com/>
- [4]. Tim Berners-Lee, "Web architecture: Metadata", <http://www.w3.org/DesignIssues/Metadata>
- [5]. Tim Bray, Jean Paoli, and C. M. Sperberg-McQueen, "Extensible Markup Language (XML) 1.0 W3C Recommendation", <http://www.w3.org/TR/REC-xml>
- [6]. Matt Richtel, "New Breed of Worker Transforms Raw Information Into Knowledge", *New York Times*, October 15, 1997.
- [7]. http://chss2.montclair.edu/prdept/_pomo/00000008.htm
- [7]. Thomas H. Davenport, et al. "Working Knowledge: How Organizations Manage What They Know", *Harvard Business School Press*, 1997
- [8]. J. Yen, C. Yang, A. S. M. Lau, A. K. L. Chung, K. Hwang, and D. Yeung, "Transparency and Efficiency in Information Transmission and Financial Digital Library", *Proceedings of the First Asia Digital Library Workshop*, pp195-206, 1998
- [9]. Ikujiro Nonaka, Patrick Reinmoeller, and Dai Senoo, "The 'ART' of Knowledge: Systems to Capitalize on Market Knowledge," *European Management Journal*, Vol 16, No. 6, pp673-684, 1998
- [10]. Narinder Singh, "Unifying heterogeneous information models: Semantic tags support knowledge webs", *Communications of the ACM*, Vol 41, No. 5, pp37-44, 1998
- [11] Hsinchun Chen, Joanne Martinez, Tobun D. Ng, and Bruce R. Schatz, "A Concept Space Approach to Addressing the Vocabulary Problem in Scientific Information Retrieval: Experiment on the Worm Community System", *Journal of the American Society for Information Science*, Vol 48, No. 1, pp17-31

From Idea to Knowledge –Generating New Knowledge on E-Business

Johanna Tommila, Anne-Mari Järvelin, Marko Seppä

eBRC / Tampere University of Technology & University of Tampere, Finland

johanna.t.tommila@uta.fi, anne-mari.jarvelin@uta.fi, marko.seppa@tut.fi

ABSTRACT

Electronization or knowledgization is under way in society, world-over. This process works the transformation of the industrial society into knowledge society, and requires a new level of border-crossing cooperation –and catalyzing of the same. This exploratory study seeks to make sense of catalyzing aimed at generating new knowledge on e-business by turning research initiatives stemming from Business, University, and Government (B.U.G.) from idea to knowledge. This action research based single case study focuses on eBRC, a university joint venture created as part of eTampere to fulfill the goals of eEurope. Theoretically, the study is anchored to network and service management literature.

Keywords: E-business, networks, service management, electronization, knowledge creation, knowledge society

1. INTRODUCTION

It can be argued that a major transformation process, of kin to industrialization, is under way in society, world-over. In the language of this study, the process referred to as ‘electronization’ or ‘knowledgization’ works to transform the industrial society into knowledge society. It can be further argued that, to generate relevant knowledge on electronization or e-business, more and deeper interaction is needed across the sectors of Business, University, and Government (B.U.G.). [9] Based on far-reaching transformation strategies that involve the entire society, ranging from local strategies, such as eTampere, to international strategies, such as eEurope, it can be concluded that new type of catalysts are needed to tackle the challenges related to knowledge generation on e-business (www.etampere.fi, www.europa.eu.int).

The mission of this exploratory study is to make sense of catalyzing aimed at generating new knowledge on e-business by turning research initiatives stemming from B.U.G. from idea to knowledge. Following an action research based single-case-study strategy, the study focuses on eBRC, a university joint venture established as part of the eTampere knowledge society program to fulfill the goals of eEurope (www.ebrc.fi). The author team includes the founding co-directors, currently director and vice-chairman of eBRC and is led by the head of eBRC’s own electronic services and presence.

eBRC is a Tampere, Finland, based virtual research center, a 50-50 joint venture between two universities, Tampere University of Technology (TUT) and University of Tampere (UTA), established for a limited life (2001-2005) as part of eTampere. The mission of eBRC is to make sense of business in the knowledge society by enabling research, increasing B.U.G. networking, building a research project portfolio, and creating new, internationally acknowledged champions in its core research fields.

Theoretically, the study is anchored to the Nordic school approach by combining network research and service management thinking [6, 22]. Deriving from this tradition the remainder of the paper is structured into two main chapters: External Analysis (networks) and Internal Analysis (service management), before summarizing the findings in Conclusions.

2. EXTERNAL ANALYSIS

The transformation from industrial society to knowledge society is marked by two megatrends or phenomena that shape the organization and the operation of different kinds of organizations: (i) rapid technological development and (ii) networking. Herein, networking refers to all kinds of inter-organizational linkages between organizations, not only those that are caused or enabled by introduction of new technologies. The external analysis of the case organization is conducted by analyzing the environmental powers by using the two phenomena as lenses through which the environment is observed. It has to be noticed, however, that the phenomena are so closely related that it is sometimes extremely difficult to differentiate between the two. Consequently, the emphasis in the following analysis is on networking.

2.1. Rapid technological development

Since the successful launch of the first commercial Internet browser, in 1995, we have witnessed a strong and rapid development of Internet related technologies. After the boom and bust, in 2000, Internet and other ICT based solutions have again steadily introduced new e-business innovations, which have had remarkable economic effects across several industrial sectors. Many traditional industrial companies have, since 2000, developed unique e-Business solutions and innovations incrementally with clear, strategy related goals with little or no remainders of the hype era [23]. The service sector is an example of an environment in which new technological innovations related to ICT-

solutions can be used in very different ways. Some services can be fully automated and transferred to e-business platforms, while other services simply cannot be automated, due to, for example, their overly 'high-touch' or complicated nature. [30]

The public sector organizations also face challenges in order to use the ICT enabled new forms and ways of operation effectively. Initiatives and practices referred to as 'eGovernment' have become an essential part of the development of public services. Public online services are perhaps the most popular ICT solution in public organizations. Such can take different forms (Internet, extranet, intranet) and cover various kinds of interactions (e.g., information exchange, negotiations, support, transactions) with different stakeholders (e.g., customers, suppliers, partners, internal customers). The concept of public online services is customer oriented in itself; the aim is to fulfill the customers' need in a cost effective way. [14, 24, 25]

eBRC's basic mission is related to enhancement of deployment –not development –of new technology. As a catalyst of electronization, eBRC aims at finding new types of business and operational models that employ the newest technology. Although development of new technology is not at core of eBRC's operation, it nevertheless creates challenges because, the faster the development, the faster new needs arise for concepts that enable deployment of the technology in question. In addition, eBRC is utilizing the newest technology in its own operation; mainly through public online service. eBRC's online service includes Internet, intranet and extranet solutions. The online service is targeted to the public, all those interested in e-business: Researchers involved in projects, research financiers, business practitioners, as well as eBRC's own team members. The different types of interactions covered include information exchange, communication and support. The special characteristic of eBRC's online service is the matching point function, which enables eBRC's key stakeholders (researchers, businesses, and research financiers) to find the right partners.

2.2. Networking

2.2.1. eBRC and its networks

It has been argued that networks of inter-organizational relationships are partly replacing traditional markets and hierarchical organizations [1, 7, 22]. Business firms and other entities enter into various kinds of dyadic arrangements such as joint ventures and strategic alliances, but also arrangements that concern numerous actors, such as supplier networks and research consortia, in order to gain strategic advantages and, simply, to operate more effectively. [29] It has also been argued that, in our rapidly changing society, networks offer a flexible way to cope with uncertainty and dynamic environmental pressures. These are

caused by globalization, technological development, regulatory uncertainties, ambiguous market signals, increasing innovation costs, unbalanced market structures, and increasing amount of information –to name the top challenges. [3, 22, 29]

In the network context, eBRC can be seen as an intermediary between different kinds of players representing researchers, businesses, and financiers. As in the case of the public sector service organizations, discussed above, eBRC also serves different customers in different levels. The services are targeted to the three groups, referred to above. In addition, eBRC serves the interests of its founder universities, as well as the underlying eTampere program, and the city of Tampere as key financier, and of course the society at large by providing the new knowledge. In traditional sense, it is also difficult to describe the suppliers of eBRC; the three main groups also represent the suppliers, as through their inputs eBRC is able to generate the value added to different customer groups. Consequently, it can be argued that the traditional view of vertical and horizontal networks (Möller-Halinen 1999) is not as such applicable to eBRC's operational format. In figure 1, the network of eBRC is described.

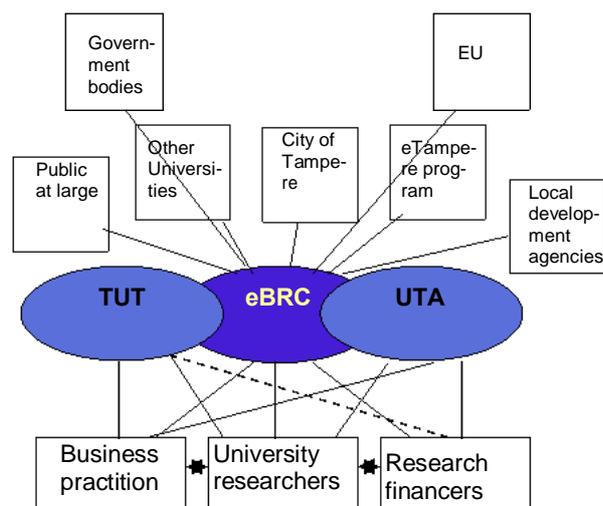


Figure 1. The eBRC network depicted

Figure 1 illustrates the network and key stakeholder groups of eBRC. As mentioned earlier, the role of the three key groups (see bottom of figure 1) can be described as both customers and suppliers, and thus the most suited nominator for them could be "partner". eBRC's relationships with its partners (partnerships) are long-term in nature, and the aim is to form mutual, cooperation based win-win relationships [1, 8]. For eBRC, the most challenging partnerships are those with businesses. During the last 20 years, Business-University relationships have gradually increased and intensified, as well as grown in relative importance. Two elements can be underscored in this development. First, universities possess such business and technical competence which, combined into a helicopter

perspective, can advance competitive advantage of business firms. Second, universities increasingly need to rely on external financing to finance their operations. Deeper Business-University cooperation has also been called for with the requirement for increased practical relevance of academic research. [28] eBRC is established by two universities with quite differing traditions with regard to Business-University cooperation. At the TUT, close relations with businesses are perceived as a natural element of scientific knowledge creation. At the UTA, such relations are clearly viewed with more skepticism, due to classic emphasis of the objectivity argument. The TUT tradition as 'the university of the industry' stems from its general roots in technology and engineering, while the UTA's more 'classic' university tradition from roots in humanities and social sciences.

eBRC's relationship with the city of Tampere is marked by the city's role as eBRC's main financier: eBRC's basic funding is based on a conditional five-year contract, in which results are evaluated twice a year. While this relationship is operatively not as deep as the relationships with the key stakeholder groups, eBRC, by definition, provides value to the city by performing its functions as a sub-program of eTampere. Of the other network partners (see top of figure 1), Other Universities form an interesting group as 'partner-competitors'. Besides cooperating, eBRC is also competing with them on research funding and business partners in research projects. These relationships can be described as 'co-optive'[2]. In such relationships the parties can cooperate in some functions and compete in others. Important is to be able to separate these two from each other in order to avoid conflicts.

2.2.2. eBRC as Intermediator

Discussion of business network intermediators has been scarce in the literature on business networks. Business networks are typically depicted as structures with no boundaries. In a network with no boundaries every actor can be regarded as intermediator as it is connected to several entities [10].

The intermediator role can also be approached as 'triad'[10, 31]. Havila [10] distinguishes between two types of triadic business relationships. First, the serial triad, is a relationship where the intermediator mediates most of the contacts between the other two actors. Serial triad reminds of a series of two dyads: (i) between intermediator and seller, and (ii) between intermediator and customer. Second, the unified triad, is a relationship where each actor has more or less equal contacts with each other, and thus this type of triad is like a group of three actors. It can be argued that, outside the context of triads, the actors can act as intermediators whose basic function is to mediate the relationships between other actors [12]. Chambers of Commerce are examples of organizations that operate

as intermediators between companies both on local and international level. eBRC can be also regarded as a specific type of intermediator in a network. The intermediating act is often performed in a certain project by, for example, finding a suitable business partner to a specific research project or helping researchers find the right financiers for their project.

In the literature related to intermediators in triads, it is argued that actors may have different roles or functions, and at different times any of the triad members may play one of the different roles. Mediator, *tertius gaudens*, and oppressor functions of the third party in the triad can be distinguished [26, 31]. The mediator is involved in both sides of the relationship (often buyer and seller), and aims at enhancing the relationship between the two by keeping them together. Thus the mediator is interested in finding solutions and creating situations, which benefit both parties. Mittilä[21] uses the term 'bistomer' for a similar type of intermediator in business relationships as mediator. The *tertius gaudens* and oppressor have the opposite attitudes as they are in the triad in order to satisfy their own interests. The *tertius gaudens* is ready to take advantage of every emerging possibility, but not taking an active role like the oppressor. [26, 31]

Järvelin and Koskela [12] have discovered two additional intermediator roles in the network context, labeled as 'initiator' and 'invisible hand'. The initiator is more or less door opener with a large contact base and a role of bringing different parties together. The invisible hand is an actor that only provides facilities for the other actors in order to help the others to build the contacts and relationships. The mediator, *tertius gaudens*, and oppressor roles can only exist in the ongoing relationships while initiator and invisible hand roles are central in the beginning of the relationships. eBRC can be in many cases regarded as an initiator which brings different actors together. In research projects more tightly connected to eBRC, it can also have a mediator role. In addition, by having created and launched an annual international research conference, eBRF (www.ebrf.fi/conference2004), eBRC also acts as an invisible hand by offering a forum for networking.

3. INTERNAL ANALYSIS

Professional service organisations (or expert services or knowledge intensive services) can be seen as true embodiments of the emerging knowledge society; the products or outputs of these organizations represent knowledge or competence as such, and can be thus regarded as being fundamental elements of the society and its development [15, 21]. eBRC's offering is new and fresh in many different aspects; the research center operates in a way that utilizes the opportunities enabled by the rapid technological and social development. eBRC thrives to be a modern, virtual, and above all

dynamic meeting place for researchers, business practitioners, and financiers – a new type of actor in the field of academic research.

In order to clarify the service and to go deeper into the concept, basic principles of service management are utilized. Professional services are being characterized as the most intangible of services. The intangibility of these services concerns not only their physical nature but also their mental characteristics. As professional services are markedly intangible, they are also often quite hard to comprehend, both for the employers, customers and other stakeholders. (11, 12, 18, 19).

The intangible nature of professional service is also present at eBRC; the concept and services offered are abstract both physically and mentally; and it has taken time to 'educate' the different stakeholder groups. The other clear professional service characteristic that is present in eBRC's operation is the almost simultaneous production and consumption (and partly also marketing) of services, which causes problems for balancing with supply and demand. This can be partly solved via 'digital presence'.

A Special characteristic of professional services is also the importance of customer participation, especially as resources as they provide valuable information, serve as references, etc. [19]. As mentioned earlier, a customer segment of eBRC also serves as its key suppliers: A prime example of customer participation in production of professional services.

3.1. Service components

Products and services are described as bundles of different benefits. Kotler's [17] three-level structure of products focuses on the core benefit which is the target of the customers' needs. This benefit is served through the actual product which includes the brand name, packaging and other features. Finally the product concept is supplemented by augmented services and benefits such as warranties. Similar models have been developed for professional services. The basic level is the core service, which defines the key problem-solving element of the service. This is the service that the customer primarily desires. Support services include conditional, additional, and surplus support services. The support services make the service accessible or give the customer more choices and options to benefit from the entire set of services. [27]

3.1.1. Core service

eBRC's core service or core benefit is project facilitation services which comprises matching of three stakeholder groups: researchers, business practitioners, and financiers. The objective is to create research and development projects by combining the stakeholders' interests. eBRC seeks to act as a catalyst in the

interface of researchers, businesses and financiers serving as a matchmaker in the process. Various research projects are conducted under the eBRC umbrella. The projects differ from expert opinions, drafted within weeks, to research programs lasting for years and comprising several subprojects. The focus is on research that is both academically valuable and purposeful for businesses.

The researchers from the founding universities serve as the key resource for eBRC and research initiatives often derive from within this group, from within each individual's own research field. An initiative for a research project can also stem from the practical needs of businesses or from the research financiers, often the EU or national government related research programs.

eBRC's role is to find and match the right partners and offer assistance in the funding process. Execution of the research projects is always led by an individual professor of a TUT or UTA institute or school and often joined by representatives of other such –outside or inside, Business or University, domestic or foreign –entities. In addition, eBRC works actively to develop different forms and tools for effective co-operation. It can also participate as a coordinator in the management of the projects.

3.1.2. Actualised service

Marketing of professional services lies usually in the hands of the professionals themselves; the service is demonstrated through earlier work. Brand-awareness is, however, increasingly important and eBRC has also developed features and services that support its brand building. eBRC's service production includes several support services. There are no particular conditional support services; the core service can be utilized as is. As an additional support service eBRC offers the option to publish research results its publication series: Research Reports (RR). The publications are available both in print and electronic format.

To augment the offering, eBRC has surplus support services. It offers its stakeholders opportunities to network with each other both virtually and physically. At eBRC's web service the visitors can, for example, discuss topical issues and acquire the latest research information. The annual conference eBRF gathers the different stakeholders together and gives an opportunity to present ongoing research for instant feedback. eBRF is multidisciplinary window on current issues related to electronization and e-business research and a border-crossing meeting ground. Participating researchers have the option to publish their final research papers in *Frontiers of e-Business Research (FeBR)*, the conference proceedings.

3.2. Electronization of eBRC

E-business in an organization can take different forms

depending upon, for example, the nature of the product, processes, or organization. The development of e-business or electronization in a specific organization evolves often through different stages. The development starts from simple Internet applications and ends up with total organizational '(e)transformation'. [5, 16]

In figure 2, the 'eState' of an organization is depicted through three dimensions: Product, process and agent [4, 30].

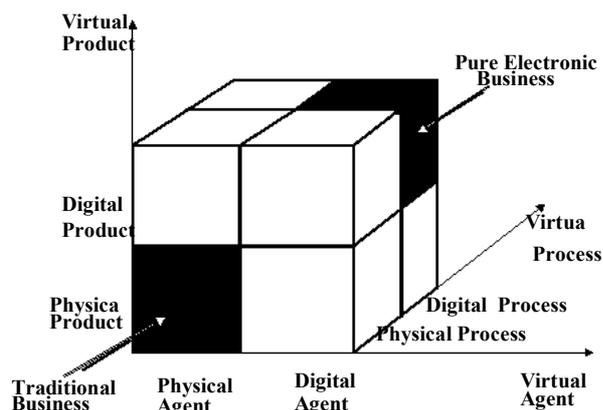


Figure 2. Dimensions of e-business

Within figure 2, eBRC would become positioned somewhere in the middle. The product of eBRC is a service, but only part of it is in digital format; the process is only partly digitalized; and the agent is still more or less physical. Consequently, in the expert or professional services, the face-to-face interactions and the personal relationships play a central role, and this also still strongly the case with eBRC as well. The professional service needs to have a face/faces in order to create the credibility needed for starting cooperation.

3.3. Service process

Another efficient way of describing the service concept is to use process descriptions. As mentioned above, services are formed through different chains of functions. Due to the project nature of eBRC's operations, each service process is individualized. However, there are also features that are similar in each process. Every project starts from an analysis of the research field, examining the different parties and their interests. The research ideas are evaluated from several points of views and the feasibility is discussed. After a project begins, eBRC's role is to facilitate the process and follow its progression. The results can be reported in eBRC's publications and possibilities for teaching implementations and further research are examined. The eBRC process is described in figure 3.



Figure 3: The eBRC process: From Idea to Knowledge

4. CONCLUSIONS

There are historic, cultural, and financing-related barriers to working closely together across the borders between universities, even individual schools and institutes within a given university, let alone those between B.U.G. Nevertheless, faced with emergent phenomena such as evidence-based-policy thinking and the increased complexity related to utilizing ICT enabled models and to understanding what is relevant to business competence, societies need to (seek to) be successful in breaking borders, not just between nations, but also sectors of economy.

While it can be argued that at least generation of new knowledge on e-business, in the emerging knowledge society, requires deeper and ever more intensive cooperation between representatives of B.U.G, this is by all means easier said than done. In this single case study, an effort has been made to introduce and analyze catalyzing of new e-business knowledge by turning research initiatives, stemming from B.U.G., from idea to knowledge. Theoretically, the exploration has utilized the Nordic school approach: A combination of network research and service management thinking.

The case organization is faced with the challenges of a typical professional service organization with regard to, in particular, (i) the dual role of key stakeholders as both suppliers and customers and (ii) electronization of its own operation. In terms of electronization, organizations and operations in the B.U.G. interface need to cope with the challenges related to ever further virtualization of their Product, Process, and –last but not least –Agent.

REFERENCES

- [1] Achrol, R. S. and Kotler, P. "Marketing in the Network Economy", *Journal of Marketing*, 63 Special Issue, 146-163, 1999
- [2] Bengtsson, M., Kock, S. "Cooperation and competition in relationships between competitors in business networks", *Journal of Business & Industrial Marketing*, 14 (3), 178-193, 1999
- [3] Cravens, D.W., Piercy, N.F. and Shipp, S.H, "New Organisational Form for Competing in Highly Dynamic Environments: The Network

- Paradigm”, *British Journal of Management*, 7 (3), 203-218, 1996
- [4] Choi, S., Stahl, D. and Whinston, A., “The Economic of Electronic commerce. The Essential Economics of Doing Business in the Electronic Marketplace”, Indianapolis: Macmillan Technical Publishing, 1997
- [5] Earl, M. J., “Evolving the E-Business”, *Business Strategy Review*, 11 (2), 33-38, 2000
- [6] Grönroos C., “From Marketing Mix to Relationship Marketing: Towards a Paradigm Shift in Marketing”, *Asia-Australia Marketing Journal* 2(1), 9-29, 1994
- [7] Gulati, R., Nohria, N. and Zaheer, A., “Strategic Networks”, *Strategic Management Journal*, 21 (3), 203-215, 2000
- [8] Halinen, A., “Exchange Relationships in Professional Services. A Study of Relationship Development in Advertising Sector”, London, Routledge, 1997
- [9] Hannula M, Järvelin A-M., and SeppäM. (eds.): “Frontiers of e-Business Research 2003”, Tampere University of Technology and University of Tampere. Tampere 2004
- [10] Havila, V., “International Business-Relationship Triads. A Study of the Changing Role of the Intermediating Actor”, Doctoral thesis no. 64. Department of Business Studies, Uppsala University, 1996
- [11] Järvelin A-M., “Evaluation of Relationship Quality in Business Relationships”, Doctoral Dissertation, Acta Universitatis Tamperensis No. 794. Tampere, 2001
- [12] Järvelin A-M. and Koskela H., “The Role of Science Parks in Developing Company Networks”, Paper published in IMP 2004 Conference Proceedings, www.impgroup.org, 2004
- [13] Järvinen R. and Järvelin A-M., “How Consumers Can Be Engaged in eInsurance?”, Mika Hannula- Anne-Mari Järvelin-Marko Seppä (eds.): Frontiers of e-Business Research 2001. Tampere University of Technology and University of Tampere. Tampere 2002. 133-146, 2002
- [14] Kannan, P. K., “Introduction to Special Issue: Marketing in the e-Channel”, *International Journal of Electronic Commerce*, 5 (3), 3-6, 2001
- [15] KemppiläS. and Mettänen P., ”Tietointensiiviset palveluyritykset – tutkimuksen nykytila (Knowledge Intensive Service Companies – the state of art)”, Sitra, Helsinki, 2004
- [16] Kettinger W. J. and Hackbarth G., “Building an E-Business Strategy”, *Information Systems Management*, 17 (3), 78-93, 2000
- [17] Kotler, P. “Principles of Marketing”, Prentice-Hall, 2004
- [18] Lehtinen U. and Järvelin, A-M., “The Abstractness of Services and Nontraditional Parameters. Applications in international services marketing”, IMDA: 3rd Annual Business Congress in Malaysia 16. - 18.6.1994., 1994
- [19] Lehtinen U. and Niinimäki S., ”Asiantuntijapalvelujen tuotteistamisen ja markkinoinnin suunnittelu (The planning of marketing and making a marketable product in professional services)”, 2004
- [20] Leiponen A., “Knowledge Services in The Innovation System” Taloustieto, Helsinki, 2001
- [21] Mittilä, T., ”Relation Trine. An Analysis of Industrial Supplier-Customer Relations”, Dissertation. Acta Universitatis Tamperensis 768, University of Tampere, School of Business Administration, Tampere, 2000
- [22] Möller, K. – Halinen A., “Business Relationships and Networks: Managerial Challenge of Network Era”, *Industrial Marketing Management*, 28, 413-427, 1999
- [23] Ruohonen M., Riihimaa J., and Mäkipää M., “From e-Business to knowledge business in metal and electronic industries”, FeBF 2002, eBRF conference Proceedings, 443-458, 2002
- [24] Rust R. and Kannan P. K., “The Era of e-Service”, in Rust R. and Kannan P. K. (eds.): e-Service: New Directions in Theory and Practice, 3-21, 2002
- [25] Rust R. and Lemon K. N., “e-Service and the Consumer”, *International Journal of Electronic Commerce*. 5 (3), 85-102, 2001
- [26] Simmel, G., “Quantitative Aspects of the Group”, A translation of Chapter II of *Soziologie*, In Wolf, K. H. (Trans., Ed.). *The Sociology of George Simmel*. Glengoe: Free Press of Glengoe, 87-177, 1950
- [27] Sipilä, Jorma, ”Asiantuntijapalvelujen markkinointi, (Marketing of Professional Services)” Helsinki, WSOY 1996
- [28] Storey, D.J., Tether, B.S., “Public policy measures to support new technology-based firms in the European Union”, *Research Policy*, 26 (9), 1037-1057, 1998
- [29] Tikkanen J. and Halinen A., “Network Approach to Strategic Management – Exploration on the Emerging Perspective”, Paper published in IMP 2003 Conference Proceedings, 2003 www.impgroup.org.
- [30] Turban E., King D., Lee J., Warketin M. and Chung H. M., “Electronic Commerce 2002: A managerial perspective”, Prentice Hall, New Jersey, 2002
- [31] Tähtinen J., Triad Net Dissolution”, Proceedings of EMAC-ANZMAC Colloquium, 16. - 17.12.2002, Perth, Australia, 2002

Knowledge Management, as the Key Factor of Survival in New Competition Age

Masoud Karami Zarandi¹, Ehsan Ghaffari Cherati²

¹Management Faculty, Tehran University, Tehran, 14335-835, Iran

²Agri & SD Corporation, Tehran, 15875-6154, Iran.

Masoud_karami_Zarandi@yahoo.com, E77ghaffari@yahoo.com

ABSTRACT

In accordance with Darwin's theory about "Survival of the fittest", whoever has better adapted himself to circumstances, will remain and continue the life, and vice versa, whoever can not do that, will die. If we consider organization as a biological organism, organization is essentially a special group that tries to survive and accord in a special environment, because organization, basically, is the thing that human has made it for the survival of himself. The biggest difference between human and other existing species that has conducted him to remain, progress, and development during many years, is the ability of thinking that others have very little quotient of that or don't have at all. History has shown that many existing species that have been apparently bigger and stronger than human, like dinosaurs, have died because of their incongruity and discordance with circumstances, but human, this apparently weak species, is continuing the life with using the essential factor of thinking and making decision. Organization contains people, and the most important factor that can help it to remain, consolidate, and surpass the competitors, in this knowledge-based economy, is thinking and using the powerful mind of organization, because nowadays, after many years from Industrial Revolution, with three production factors of land, capital and labor, in the century of information and creating organized R&D, knowledge and management have become the most important factors of production, and those organizations which better use these two factors, are more successful and permanent. Thus creating and then continues improvement of knowledge management in organization, as a process that create the "mind of organization", is inevitable, and during this course, not only knowledge of the organization's members will be gather and according to the role of the synergy, total knowledge of a group is more than all their individual knowledge, but also available data and information from environment, specially from customers, would constantly enter as the fuel of organization's mind and lead to "develop" the mind of organization" which the result of that is performance found on the internal and external circumstances of organization, and finally "survival of the organization as the fittest". This paper is going to consider and survey creating this principle of the survival of organization (the mind of organization) and improve that, technologically and socially.

Keywords: Survival of the fittest, Knowledge management, Organizational memory, Organizational mind.

1. INTRODUCTION

In accordance with Darwin's theory about "survival of the fittest", whoever has better adapted himself to circumstances, will remain and continue the life, and vice versa, whoever can not do that, will die. Because due to resource limitation, there has been a continues competition between different existing species for survival. As Henderson has explained in his paper (The origin of strategy), competition began with life itself. The first one-cell organization required certain resources to maintain life. When these resources were adequate, the number grew from one generation to the next. As life evolved, these organisms became a resource for more

Complex forms of life, and so on up the food chain (Henderson, 1989). Human being may be at the top of the ecological chain, but we are still members of the ecological community, and therefore, no aspect of human life is untouched by Darwin's theory of evolution-survival of the fittest (Lessem, 1998).

Darwin's theory has shown, in the best way, how some existing species have been remained and how, some of them like dinosaurs, in spite of being bigger and stronger, have died. When any pair of species competed

for some essential resource, sooner or later one displaced the other (Henderson, 1989). In the other word, the fittest survive and prosper until they displaced their competitors or outgrow their resources (Henderson, 1989). Today more than a million distinct existing species has been cataloged, each with some unique advantage in competing for the resources it requires. At any given time, thousands of species are becoming extinct and thousands more are emerging (Henderson, 1989). The species that remain in this competition, have adapted themselves to environment by different unique ways, and human, as a kind of species, since a few millions years ago has also dealt with this competition. Living in an unstable and ever-changing environment has been his challenge and has constantly struggled for remaining and survival. History shows that the challenging way of survival has not been always the same, and has been changed in different times. Considering, Alvin Toffler's segmentation, as a base of this analyses, we can find that, human was exactly living like animals in "the age of Hunting", he has been ever in chasing and escaping. He did not have any fixed place and was continuously looking for other places for better hunting. But reviewing the archeological findings shows that, human; even at those starting days, has used from his mind, although simply. After that, with

emerging “the age of agriculture” and finding the importance of land, human has been gradually depended to the land, because he found to improve and develop his life and accord with circumstances that have surrounded him, again using the key factor of MIND. Furthermore, in this age, human, depending on the land and changing from wandering to dwelling, has gradually found the importance of co-operation and so has created the first step of emerging organization as a group of a people. In the next age, creating vapor machine and then Industrial revolution, has reduced the importance of land, as the most important factor of production, and has focused on capital, and whoever had more capital could better compete and remain in the field of competition. Big factories were created in this age and started mass production and distribution of uniform products in markets. After that emerging scientific management, has increased the importance of thinking, decision making and planning and it gradually lead to another evolution in human’s life, with using the power of MIND and thinking.

Inventing computer, developed the use of data, and prepared a field in order to gathering, analyzing, storing and applying it, and then creating internet, led to a limitless and high speed communication, with low cost, and all of these events caused to a basic change over, that reduced the role of traditional production factors and increased the importance of data as the base of businesses. In new age that is called “the age of network and digital economy” or “the age of information”, nobody and no organization is not able to be successful individually. Borderless and high speed communication, link organization and their customers together, and therefore survival of each, strongly depends to another.

In this new age that philosophers called postmodernism, data, information, and knowledge are playing the biggest role in the field of competition, because according to the postmodernism, market is customer's mind, and therefore it is not possible to obtain this post modern's complicated and ever-changing market, without proper knowledge. The postmodern consumer takes on multiple, sometimes even contradictory projects, to which s/he is marginally and momentarily committed, not taking anyone too seriously. This loose of commitment is observed in all walks of life. The consumer of postmodern culture appreciates and enjoys the paradox and the playfulness, the difference and the stair that such juxtapositions provide and enable. Contemporary actions of consumers tend to indicate that they maybe more "fickle" than explainable and, therefore, predictable. Consumers frequently change their self-concepts, characters, values and etc (Firat, Dholakia, Venkatesh, 1994).

So appropriate using of data, as the base of information and knowledge, is the most critical factor of the survival of organization, and nowadays, as a matter of fact,

gathering and optimum using of data is the biggest fear and apprehension of organizations.

As mentioned, the concept of organization has been strongly changed during different ages (since the age of agriculture up to now), and today, structure of the organization is quite different.

2. ORGANIZATION AS A WAY TO SURVIVAL

The great social transformations in history have been essentially organizationally based. The Roman Empire, the spread of Christianity, the growth and development of Capitalism and Socialism, and the current changes taking place in both Capitalism and Socialism have been and are accomplished through organizations (Hall,1991).

Depending on the land, and finding the importance of living together and co-operation, created the first step of emerging the organization during the age of agriculture, and this course has developed and progressed up to now. Organizations are made of people and their relationships with one another (Daft, 1998).

This paper emphasizes that emergence of the organization, is essentially an attempt to the survival of human, which is not able to compete with competitors, individually. So, organization for human is a devise to accord him better with environment. The organizations are, fundamentally, social groups attempting to adapt and survive in their particular circumstances. Thus, formal organizations, like all other social groups, are governed by one overriding goal: survival (Scott, 1987). Thus organization strives to survive and to maintaining its equilibrium. And as Morgan says, organizations are open systems that need careful management to satisfy and balance internal needs and to adapt to environmental circumstances (Morgan , 1997).

The instinct of the survival of organization is due to some people who are beneficiary in organization, because organization is the source of power, wealth, prosperity, or dignity for them, so the desire of this group of the people, for keeping organization is usual, and in fact, they can help themselves to remain and progress, by keeping and developing organization. But this organization that is made in order to better accordance and finally survival, has some continues challenges in dealing with environment and competition. Hennan and Freeman in their article “The population economy of organization” have mentioned that, there are a number of obvious limitations on the ability of organizations to adapt. That is, there are a number of processes that generate structural inertia. The stronger the pressure, the lower the organizations’ adaptive flexibility and more likely that the logic of environmental selection is appropriated. In each distinguishable environmental configuration one finds, in equilibrium, only that organizational form optimally

adapted to the demands of the environment. In the other word, selection eliminates the weakest competitors (Shafritz, Steven ott, 1992).

So organization, like the other species in contact with environment and circumstances, accord and survive, or will be force to defeat and passing life. The environment selects out optimal combinations of organization (Shafritz, Steven ott, 1992). When any pair of species competed for some essential resource, sooner or latter one displaced the other (Henderson, 1989). As long as the resources with sustain organization are finite and populations have unlimited capacity to expand, competition must ensure (Shafritz, Steven ott, 1992). If every business could grow indefinitely, the total market would grow to an infinite size on a finite earth. It has never happened. Competitors perpetually crowd each other out (Henderson, 1989). Therefore, the organization that human has created for the survival, requires finding an appropriate solution for survival, in the complicated field of competition with other organization. This competition has strongly had a jumpy growth in the new age and has dramatically increased the speed of birth and death of organization. So that nowadays few organizations experience third decade of their life.

3. NEW COMPETITION AGE

The history of human's life, achieving to the land, gold and oil, has been the source of power and wealth. But suddenly, it was changed and knowledge was replaced. As a mater of fact, knowledge is the foundation stone of wealth (based on Lester Thurow,, 2000).

Today, the competitiveness of the firm relies less on traditional factors (capital, land and labor) than was true in the past (Sher and Lee, 2004). Moreover, knowledge will become not just a source of competitive advantage, but also the only source of it (Drucker, 1993).

In the twenty-first century landscape, firms must compete in a complex and challenging context that is being transformed by many factors, from globalization, technological development, and increasing rapid diffusion of new technology, to the development and use of knowledge (Hitt, Keats and Marie, 1998). And as Marquardte says, We live in the global age (Marquardte, 1999). This new landscape requiring firms to do things differently in order to survive and prosper. Specifically, they must look to new sources of competitive advantage and engage in new form of competition (Jackson, Hitt, Denisi, 2003).

Therefore, organization was essentially faced to a complicated competition, because the priority of ever-changing data, information and knowledge. Peter Drucker says that: Increasingly, successful organizations are building competitive advantage through less controlling and more learning- that is,

through continually creating and sharing new knowledge (Drucker, etal. 1997).

And we emphasize that; knowledge management is the best way to adapting circumstances and achieving competitive advantage.

Most organizations today are focused on the value of knowledge (Plessis and Boon, 2004). Knowledge management equips organization to the same thing that was the factor of human's survival, during a few million years ago, and this determinant and effective factor is "Mind".

4. KNOWLEDGE MANAGEMENT

As we have mentioned above, today, knowledge has been becoming the main source of the organization, moreover knowledge will become not just a source of competitive advantage, but also the only source of it (Drucker, 1993).

Current management interests are also focused on knowledge management as a major determinant of business excellence and competitive advantage (Sher and Lee, 2004).

Primarily knowledge is about the way human being has constantly evolved and learnt from his experience. It is the embodiment of this soul experience; it is about his ability to analyzing it and passing on to the next generation. Even in the corporate context the sentiments may not change (Raman, 2003).

Organization, as a group of people, is consist of different individual minds and we constantly believe that an organization that wants to remain and survive in this knowledge-based environment, must equip himself to an integrated MIND. Knowledge management, as a powerful competitive weapon (Chang, 2004), is exactly the same thing that can lead to emerging a sustained competitive advantage in the complicated field of competition, with elaborate configuring of organization's MIND. In the other word, knowledge management is the key distinguishing factor of successful firms (Lee and Cho, 2003), and has also been described for its possible role in creating sustained competitive advantage for organization (Grant, 1996; Johansson and Olsen, 2003; and Lad and Wilson, 1994). Likewise, firms with high knowledge management capability in a key area should be able to respond very quickly to strategic moves by competitors (Gold etal, 2001), in order to gain competitive advantage over competitors (Grossman and Packer, 1998).

Knowledge management systems must enable knowledge to flow efficiently in order to enhance productivity, quality, innovation and business excellence (Sher and Lee, 2004). Also knowledge management reduces the operating costs of a firm and

creates added value to customers by significantly increasing product quality (Ofek and Sarvary, 2001). When the knowledge assets of the firm are exploited, the firm sees enhanced dynamic capabilities and increased business value. And dynamic capabilities refer to an organization's way of responding in a rapidly changing environment (Helfat, 1997).

Many researchers have defined knowledge management from various perspectives. For example, Raman explains that knowledge management is the way to keep knowledge growing through sharing (Raman, 2003).

In the other definition, knowledge management has been defined as an effort to not only explicit factual information, but also the tacit information, a knowledge that exist in an organization, usually in the minds of employees in order to advance the organization's mission (Broadbent, 1998; Davenport and Prusak, 1998). In the other word, knowledge management involves blending a company's internal and external information and turning it in to actionable knowledge via a technology platform (Dimattia, 1997).

This paper defines knowledge management as the process of creating appropriate solutions in accordance with circumstances, through social systems of organization that supported by technical systems. In fact, organization creates the process of learning, with using knowledge management. Learning has been defined as the process of acquiring new information (Braun, 1998). So organization, with using this process, is continuously exchanging data to information and knowledge, and finally achieving to a collective knowledge. From this perspective, learning is never complete, as new data, new events, or new contexts create opportunities for additional improvement, insight and innovation (Lessem, 1998). Continuing this process of learning, lead to a collective memory. By definition, memory refers to the persistence of learning in a state that can be revealed at a later time (Braun, 1998). A form of collective memory is organizational memory (Paul, Haseman and Ramamurthy, 2004).

Many researchers have defined organizational memory from various perspectives. For example, Stein and Zwass, defined organization memory as means by which knowledge from the past is brought to bear on present activities, thus resulting in higher or lower levels of organizational competitiveness (Stein and Zwass, 1995). And Chang, Cho and Lee have described it as a computer supported human interactive mechanism to store and facilitate knowledge for enhancing organizational effectiveness (Chang, Choi and Lee, 2004). So, organizational memory also can help organization to respond appropriately to environment with keeping and directing knowledge, but knowledge management is not only limited to acquiring, storing and exchanging the information, but also require to interpretate and organize information from various

perspectives. So, what ever that organization needs for survival, is more than memory, and this missing link of organization is MIND. Processing data can be performed by machine, but only the human mind can process knowledge or even information (Shera, 1983). Mind is the ability to be aware of things and to think and feel, or in the other word, mind is the source of one's thoughts and feelings. And Organizational mind in this paper has been described as the ability of team thinking and analyzing internal and external environment of organization, that enable the organization to feel properly and respond effectively to ever-changing and unstable circumstances better than competitors. If we think so about the organization, every aspect of organizational functioning depends on information processing of one form or another. In our view, organizational mind relies upon the interaction between individuals and technologies, so that it is not imitable because of the unique complex of individual and their interactions in each organization. In fact, in any competitive landscape it has been argued that intangible resources are more likely to produce a competitive advantage because they often are truly rare and can be more difficult for competitors to imitate (Jackson, Hitt, Denisi, 2003).

At this regard, systems (organizations) must have the capacity to sense, monitor, and scan significant aspects of their environment. They must be able to relate this information to the operating norms that guide system behavior. They must be able to detect significant deviations from these norms and, must be able to initiate corrective action when discrepancies are detected. If these four conditions are satisfied, a continuous process of information exchange is created between system and its environment, allowing the system to monitor changes and initiate appropriate response (Morgan, 1997).

5. CONCLUSION

This paper has been tried, with considering Darwin's theory regarding to survival of the fittest and dieing those existing species that are not able to adapt themselves to circumstance, has tried to show that human as a kind of existing species that has been placed at the highest point of biological organism chain, is not an exception, and has variously attempted to be accord with circumstances, all over his life, during different ages. But the thing that has helped him during these long periods, was been distinguishing factor of MIND.

With referring to Tofler's segmentation about the route of civilization, we have mentioned, how human can adapt himself with circumstances in each period, and with a comprehensive discussion of new age that has been called the age of information and digital, we have attempted to state unique specifications of the new age.

In the other section of this paper we have described the course of emerging and transformation of organization as a device that human has created in order to the survival. Knowledge management in this knowledge-driven age is the only one thing that can lead to a unique competitive advantage with conducting some interaction between individuals, technologies and methods. Therefore organizational mind is the missing link of organization, and the key distinguishing factor of survival, in new competition age. And finally, as Winston Churchill has truthfully said several years ago "the empires of the future are empire of the mind".

REFERENCES

- [1] braun, Kathryn, "Mind of the market: constructive memory process", *Harvard Business Review*, Vol. No. Pp1-4, 1998.
- [2] Broadbent, M, "The phenomenon of knowledge management: what does it mean to the information professions?", *Information Outlook*, Vol. 2, No.5, Pp23-35, 1998.
- [3] Chang, Jaegyong, Byounggu Choi, and Heesok Lee, "An organizational memory for facilitating knowledge: an application to e-business architecture", *Expert Systems With Applications*, Vol.26, No. 2, Pp203-215, 2004
- [4] Chuang, Shu, Hui, "A resource-based perspective on knowledge management, capability and competitive advantage: an empirical investigation", *Expert System With Applications*, Vol. 27, No. 3, Pp459-465, 2004
- [5] Chang, Jaegyong, Byounggu Choi, and Heesok Lee, "An organizational memory for facilitating knowledge: an application to e-business architecture", *Expert systems With Applications*, Vol. 26, No.2, Pp203-215, 2004
- [6] Drucker, P.F. , *The Post-capital Society*, New York: Harp Collins, 1993.
- [7] Drucker, P.F., E. Dyson, C. Handy, P. Saffo and P. Senge, *Looking ahead: implications of the present*, Harvard Business Review, 1997.
- [8] Daft, Richard, L. , *Organization Theory And Design*, Ohio: South Western Publishing, 1998.
- [9] Dimattia, S., Oder, N., " Knowledge management: Hop, hype or harbinger?", *Library Journal*, Vol. 122, No. 3, Pp33-35, 1997.
- [10] Fuat, Firat, A., Nikhilesh Dholakia, Alladi Venkatesh, "Marketing in a postmodern world", *European Journal Of Marketing*, Vol. 29, No.1, Pp40-56, 1995.
- [11] Gold, A.H., A. Malhotra and A.H. Segars, " Knowledge management: an organizational capabilities perspective", *Journal Of Management Information System*, Vol. 18, No. 1, 2001.
- [12] Grant, R.M." Prospering in dynamically competitive environments: Organizational capability as knowledge integration", *Organizational Science*, Vol.7, No. 4, Pp375-387, 1996.
- [13] Grossman, R.B., "Betting the business: strategic programs to rebuild core information systems", *Office, Technology and People*, Vo.5, No.4, Pp235-243, 1989.
- [14] Henderson, Bruce, D., " The origin of strategy", *Harvard Business Review*, Pp139- 143, 1989. 5
- [15] Helfat, C.E., " Know-how and asset complementarity and dynamic capability accumulation: The case of R&D", *Strategic Management Journal*, Vol.18 , Pp339-360, 1997.
- [16] Ho, Cheng-Ter, Yuh-Min Chen, Yuh-Jen Chen, Chin-Bin Wang, " Development a distributed knowledge model for knowledge management in collaborative development and implementation of an enterprise system", *Robotics And Computer-Integrated Manufacturing*, Vol.20, No.5, Pp439-456, 2004
- [17] Hitt, M.A., B.A. Keats and S.M. Demarise, " Navigating in the new competitive landscape: Building strategic flexibility and competitive advantage in the 21st century", *Academy Of Management Executive*, Vol. 12, No. , Pp22-44, 1998.
- [18] Hall, Richard, H., *Organizations: Structures, Processes And Outcomes*, New Jersey: Prentice-Hall International Inc., 1991.
- [19] Johnsonessen, J.B., Olsen, " Knowledge management and sustainable competitive advantage: The impact of dynamic contextual training", *International Journal of Information Management*, Vol. 23, Pp277-289, 2003.
- [20] Jackson, Susan, E., Michael, A. Hitt, Angelo, S. Denisi, *Managing Knowledge For Sustained Competitive Advantage*, New York: John Wiley & Sons, Inc., 2003.
- [21] Lado, .a., M.C. Wilson, " Human resource systems and sustained competitive advantage: a competency-based perspective", *Academy Of Management Review*, Vol.19, No. 4, Pp699-727, 1994.
- [22] Lessem, Ronnie, *Management Development through Cultural Diversity*, New York: Rotledge, 1998.
- [23] Marquardt, Michael, J., *The Global Advantage: How World-class Organizations Improve Performance Through Globalization*, New York: Gulf Publication Company, 1999.
- [24] Morgan, Gareth, *Images Of Organization*, California: SAGE Publication LTD., 1997.
- [25] Ofek, E., M. Sarvary, "Leveraging the customer base: creating competitive advantage through knowledge management", *Management Science*, Pp1441-1456, 2001.
- [26] Paul, Souren, William D Haseman and K. Ramamurthy, " Collective memory support and cognitive-conflict group decision making: an experimental investigation", *Decision Support System*, Vol.36, No.3, 2004
- [27] Plessis, M.DU., J.A. Boon, " Knowledge management in e-business and customer relationship

- management: South African case study findings", *International Journal Of International Journal Of Information Management*, Vol.24, No.1, Pp73-86, 2004 6
- [29] Raman, Athothathri, *Knowledge Management*, New Dehli, Excel Books, 2003.
- [30] Scott, Richard, W., *Organizations: Rational, Natural, And Open Systems*, New Jersey: Prentice-Hall, Inc. 1987.
- [31] Shafritz, Jay M., J. Steven ott, *Classics Of Organizational Theory*, California: Brooks/Cole Publication Company, 1992.
- [32] Stein, E.W., V. Zwass," Actualizing organizational memory with information systems", *Information System Reaserch*, Vol.6, No.2, Pp85-117, 1995.
- [33] Shera, J. liberarianship and information science, in Machlup,F., and Mansfield,U.(eds), *The study of information: Interdiciplinary message*, New York: John Willey& Sons, 1983.
- [34] Sher, Peter, J., Vivid, C. Lee," I nformation technology as a facilitator for enhancing dynamic capabilities through knowledge management", *Information Management*, Vol. 41, No. 8, Pp933-945, 2004

Study on the Cultivation of Core Competence Based on Tacit Knowledge

Juanru Wang, Songzheng Zhao, Jin Yang

School of Management, Northwestern Polytechnical University, Xi'an 710072, China
wjuanru@163.com, jinyang1973@163.com, zhaosongzheng@263.net

ABSTRACT

This paper discusses the meaning and characteristics of core competence, analyzes the relationships between tacit knowledge and core competence, and gets the conclusion that tacit knowledge is the basis for the development of core competence. At last methods that can transfer tacit knowledge into core competence are discussed.

Keywords : Core Competence, Tacit Knowledge, Cultivation

1. INTRODUCTION

In the new business environment, characterized by an increased turbulence and complexity, an organization's capacity to create and sustain competitive advantages lies in what it knows, not what it owns. Therefore, knowledge, especially tacit knowledge, is increasingly considered an important source of sustainable competitive advantage. Thus, it is necessary for tacit knowledge to be converted into competence. So it is important and meaningful to study the cultivation on how to transfer tacit knowledge into core competence.

2. THE CONCEPT OF CORE COMPETENCE

C.K. Prahalad and Garry Hamel mentioned core competence in the paper of enterprise's core competence published in the Harvard Management comment. They think enterprise's core competence is accumulated knowledge in the organization, especially the knowledge that coordinates different product skills and combines with many technological schools^{[1][2]}. It is a set of special skills and technologies that can provide additional value to customers. Dorothy Leonard-Barton defined enterprise's core competence as knowledge system identifying and offering advantages. It consists of four elements, knowledge base and skill, technical system, management system, and value system. Thus, core competence is a knowledge system embedded by enterprise's tacit knowledge, and must be accumulated by enterprise^[3]. It is valuable and unique for an enterprise, which is difficult to be imitated and acquired by competitors.

From the concept of core competence, its characteristics can be summarized: 1) It is unique, which mainly manifests in tacit knowledge that is exclusive for enterprise and unique for whole industry^[4]. 2) It is customer-oriented value, that is, it can make enterprise provide more value to customers. 3) It is dynamic. Knowledge structure, storage, and capability of enterprise will change with time and external environment, so core competence must be continuously updated. 4) It is synthetic. Core competence combines with many kinds of knowledge and technologies, and it

touches every aspects of enterprise. 5) It is inimitable. Once it forms, core competence will become the unique competitive advantages of enterprise, and other enterprises can not acquire this capability in short-term. 6) It has path-dependence. The formation of core competence depends on synthesizing existing knowledge storage, knowledge structure, and enterprise's absorptive capability, which all have stronger path-dependence.

3. RELATIONSHIPS BETWEEN TACIT KNOWLEDGE AND CORE COMPETENCE

Knowledge can be classified as tacit knowledge and explicit knowledge from the point of view epistemological. Explicit knowledge is defined as encoded knowledge, which can be expressed by words and figures, so it is easy to communicate and imitate by competitors. Tacit knowledge, however, is neither easily imitated nor clearly understood outside the enterprise, i.e. skills and practical knowledge. It can be classified into four types: scientific, conscious, communal, and automatic^[5]. Scientific knowledge is social, abstract, and highly mobile. In this case, appropriability is an important issue. Alternatively, appropriability is less of a concern with automatic (individual-focused and taken for granted) or conscious (individual-focused and explicit) knowledge. To use automatic or conscious knowledge effectively, it must be integrated with the enterprise's complementary assets. Moral hazard is of concern if an individual or teams holds the tacit knowledge. Communal knowledge, on the other hand, is often interrelated with the culture. Four types of tacit knowledge are all enterprise specific and often can not be written or encoded. They also have an immutable, hard-to-decipher quality that can not be easily transmitted to others, and often represent a shared experience among organization members. They not only can be transferable within the enterprises relatively and widely, but also can be very difficult for other enterprises to copy or develop. So tacit knowledge is the basic resources for cultivating enterprise's core competencies.

The difficulty of expressing, codifying, and transmitting

tacit knowledge makes it easier for an enterprise to protect than that of explicit knowledge. So acquiring another enterprise's tacit knowledge requires hiring its people. But a single expert is unlikely to possess all of the knowledge needed to reestablish the desired organization in the new enterprise, since the knowledge underlying core competence is held by a number of people whose individual knowledge is incomplete. Moreover, tacit knowledge may only be effective when embedded in a particular enterprise culture, structure, a set of processes and routines. The difficulty of copying tacit knowledge enables tacit knowledge to be the basis of an inimitable competitive advantage, and usually, tacit knowledge can be converted into core competence only through transferring^[6], as show in Figure 1. For

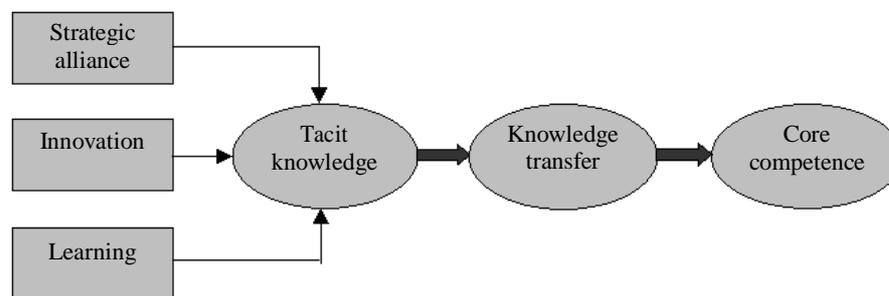


Figure 1 Relationship between tacit knowledge and core competence

4. APPROACHES OF CULTIVATING CORE COMPETENCE

Because tacit knowledge is much difficult for competitors to imitate than that of explicit knowledge, the ability to capture and transfer tacit knowledge is the key for cultivating competitive advantage. Following is some methods to spread tacit knowledge within an enterprise so that it can be transferred core competence.

4.1 Learning by Apprenticeship

Learning is often the route by which many enterprises can accelerate development of a knowledge base. Explicit knowledge can be readily transmitted and learned by others through documents. Tacit knowledge, on the other hand, is often captured in metaphors that outsiders can not easily decode, so it often can only be learned by apprenticeship or learning-by-doing.

Learning by apprenticeship is a good method to translate tacit knowledge, it helps the enterprise understand and define complex problems and, therefore, solve them in unique way^[7]. Several things can be done to promote learning by apprenticeship in companies. First, managers can be taught how to coach. Second, their performance measures can include their efforts to train subordinates. Third, structured group discussions can be used to serve many of the educational functions of training with far lower costs in terms of human resources than that of one-to-one mentoring. For

example, the study of American government indicates the cost of each auto made in Japan is less 2200 \$ than that in America, because Japanese auto manufacturing uses advanced inventory control, personnel management, and quality management. So many large-scaled enterprises in American, such as, Ford, GE, and Chrysler, put many resources to mine and study the product knowledge of Japanese auto. Although great deals of data were collected, many consultants were invited, and relative books were written, American enterprises did not acquire competitive advantages like Japanese enterprise. Obviously, the successful advantage of Japanese enterprise based tacit knowledge is difficult to be imitated by others.

example, in enterprise, R&D department hold regular conferences in which difficult problems are shared and discussed. Each participant has the opportunity to assess how the problems can be dealt with, and to hear the professional members of the department how to deal with the problems.

4.2 Making effective communication

Tacit knowledge can be transferred through personnel communication. The effective communication mechanisms in enterprise are net and teamwork. In enterprise, people can work through net informally with others having similar interests, or form communities of practice, to discuss their experiences, gather the ideas of others, and receive feedback on their own ideas. Communities of practice are groups of people sharing an interest in an issue who meet periodically to discuss problems, brainstorm, and share knowledge^[8]. Meetings may be in person, or virtual. Optimally, information is stored in a database for future use. Companies can significantly facilitate the transmission and spread of tacit knowledge by supporting such net: providing the communications resources needed, giving people time to use net, and formally recognizing the community and its best contributors. When supported by the enterprise, communities of practice hold considerable potential for developing and spreading the expertise of an enterprise.

A common problem in teamwork is that they tend to deal with explicit knowledge, rather than tacit

knowledge. Therefore, companies need to arrange case discussions and other opportunities for individuals to exercise judgment and compare their decisions with those of experts. Working in teams can also present a platform for sharing tacit knowledge. When working in teams, people have an opportunity to observe how others conceptualize situations, approach problems, generate and evaluate solutions. The more people work together, and the more time they spend on socializing and talking about their experiences and sharing anecdotes, the more tacit knowledge they will share.

4.3 Fostering Dynamic Routines

Another approach of cultivation core competences is instilling dynamic routines within the organization. Organization's dynamic routines fostering differentiated technological capabilities represented a critical enterprise-specific resource. Broadly speaking, dynamic routines refer to the organization's particular approaches that provide the basis for understanding and transferring tacit knowledge^[5]. Routines are important for influencing the context of corporate strategy. Enterprise-specific routines established within the organization are based on the collective consciousness among its members. The complexity and dynamics of the organization's routines play a critical role in transferring tacit knowledge in the enterprise.

Moreover, numerous organizational routines enable the enterprise to envision and to develop combinations of different tacit knowledge that allow for more competitive. Combinations of routines allow for enterprise-specific combinations of tacit knowledge that are applicable to numerous products and processes. As a result, the risk of entering new markets is reduced. Accumulation of dynamic routines generates scaled economic benefits in learning. Dynamic organizational routines reflect the richness of the enterprise's learning patterns, as well as its potential for developing idiosyncratic know-how. And enterprise's ability to absorb new knowledge and to create new knowledge is related to the diversity of its learning base. Particularly in uncertain environments, the promotion of several organization routines enables the enterprise to absorb different types of knowledge, especially tacit knowledge to develop new skills faster. Therefore, application of dynamic routines increases causal ambiguity for competitors and increases the attractiveness of entering new markets.

5. CONCLUSIONS

Tacit knowledge is the basis for cultivation competitive advantages. It can be spread within an enterprise, but it is very difficult for other enterprises to imitate. The predominant way that enterprises can facilitate the spread of tacit knowledge within themselves is by giving employees the chance to work with expertise. In addition, making effective communication and fostering dynamic routines also provide employees with the opportunity to develop their own tacit knowledge and communicate with other. The more tacit knowledge are communicated and transferred in enterprise, the easier core competence develops.

REFERENCES

- [1] Nigel Petts, "Building Growth on Core Competences---a Practical Approach", *Long Range Planning*, Vol. 30, No. 4, pp551-561, 1997.
- [2] Wei Jiang, "Carrier and Characteristics of Knowledge-Based Core Competence", *Science Research Management*, Vol. 20, No. 2, pp55-60, 1999.
- [3] Li Huanrong, Linjian, "The cultivation of Enterprise's Core Competence Based on Knowledge Alliances", *Science Research Management*, Vol. 22, No. 3, pp20-26, 2001.
- [4] Olivier Boiral, "Tacit Knowledge and Environmental Management", *Long Range Planning*, Vol. 35, No. 3, pp291-317, 2002.
- [5] Jon-Arild Johannessena, Bjørn Olsen, "Knowledge management and sustainable competitive advantages: The impact of dynamic contextual training", *International Journal of Information Management*, Vol. 23, No. 4, pp277-289, 2003.
- [6] Liao Zhenpeng, "Knowledge, Innovation and Core Competence", *Journal of Kunming University of Science and Technology*, Vol. 27, No. 4, pp153-157, 2002.
- [7] Roy Lubit, "Tacit Knowledge and Knowledge Management: The Keys to Sustainable Competitive Advantage", *Organizational Dynamics*, Vol. 29, No. 4, pp164-178, 2001.
- [8] Kaj U., Koskinen, Pekka Pihlan, Hannu Vanharant, "Tacit knowledge acquisition and sharing in a project work context", *International Journal of Project Management*, Vol. 21, No. 4, pp281-290, 2003.

The Relationship between Knowledge Sharing Behavior and the Effectiveness of Information Systems Strategic Planning (ISSP): An Empirical Study in Taiwan

Jungchi Pai

Department of Business Administration, Takming College, Taipei, Taiwan, China
E-mail: rogerpai@mail.takming.edu.tw

ABSTRACT

“Improving information systems strategic planning” remains among one of top ten issues facing IS/IT executives and corporate general managers. As e-business strategies have received growing attention from enterprises, information systems strategic planning (ISSP) is now considered critical in developing a successful e-strategy. However, numerous researchers have examined the relationship between various factors and the effectiveness of information systems strategic planning, the effects of knowledge sharing behavior on the effectiveness of information systems strategic planning still have little been examined. The main purpose of this study is to examine the relationships between knowledge sharing and the effectiveness of information systems strategic planning. Furthermore, we explore the factors influencing knowledge sharing behavior among stakeholders in the ISSP process. Data were collected by a questionnaire survey sent to the IS directors of 805 large companies in Taiwan. Survey results indicate that knowledge sharing behavior influence of the effectiveness of information systems strategic planning. The implications of this finding for practice and future research are also examined herein.

Keywords: Information systems strategic planning (ISSP), e-business strategy, knowledge sharing

1. INTRODUCTION

ISSP has been identified as essential in integrating IT into an organization to increase a firm's strategic competitive advantage (Huysman *et al.*, 1994; Lederer and Sethi, 1996; Mentzas, 1997; Levy *et al.*, 1999). With the recent growth in interest in e-business and e-commerce, information systems strategic planning (ISSP) is widely viewed as an effective means of implementing a successful e-strategy (Lee and Pai, 2003). As the information technology environments and information systems applications are growing complex, the ISSP process that cannot be only handled by information systems professionals. A strategic IS/IT planning team which comprised of business managers, IS managers, user managers and unit managers can help the organization to achieve consensus through effective communication and interaction and achieve their information systems decision-making goals. Since these stakeholder groups have specific different explicit knowledge and implicit knowledge, how to effectively manage the knowledge and enhance knowledge sharing among stakeholders to be one of the most important issues in the ISSP process. Four types of knowledge must be integrated in the ISSP process, including business knowledge, organization-specific knowledge, IT/IS knowledge and management competencies. Furthermore, much research has been conducted in the area of ISSP in recent years. However, the relationship between knowledge sharing behavior among stakeholders and information systems strategic planning has been seldom examined empirically.

The main purpose of this study is to examine the

relationship between knowledge sharing behavior and the effectiveness of information systems strategic planning. Moreover, this study also examine the factors affecting the knowledge sharing among stakeholders in the IS strategic planning process. The hypothesized relationships were empirically tested using a field survey of Taiwan's large firms.

2. THEORETICAL BACKGROUND

Information systems strategic planning has been described as a managerial and interactive learning process for integrating information systems considerations into the corporate planning process, aligning the application of information systems to business goals, developing detailed information systems plans and determining information requirements to achieve business objectives (Earl, 1989; Galliers, 1991; Teo and King, 1997; Cunningham, 2001). A growing amount of empirical work that has been conducted examining the relationship between ISSP and organizational context. However, there has been little empirical research that investigates the effects of knowledge sharing behavior among stakeholders on the effectiveness of information systems strategic planning. A lot of various types of knowledge must be integrated in the ISSP process such as IT/IS knowledge and business knowledge. Such knowledge may be “tacit” or “explicit” (Nonaka, 1994). Tacit knowledge resides in the minds of different stakeholders and has not been documented in a structured form, but such knowledge is important to the planning process. For example, the CIO possesses IT knowledge and experience, the CFO possesses organization-specific knowledge concerning finance and accounting, and CEO possesses business executive

knowledge and experience. These senior executives accumulate their IS/IT strategic planning knowledge over time through interactive learning processes. Different varieties of tacit knowledge exist in the minds of different interest groups, all with their own expertise, positions, perceptions, and powers. Consequently, these differences may result in difficulties in cooperation, and may affect the effectiveness of ISSP. To achieve the consistently planning objectives, “knowledge sharing” is necessary to the ISSP process.

3. RESEARCH MODEL AND HYPOTHESES

Based on the literature reviews, the research model (Figure 1) was first built, using knowledge management, behavioral, and IS planning theory as reference disciplines. The research model has three major parts, including knowledge sharing behavior, the factors affecting the knowledge sharing behavior and the effectiveness of information systems strategic planning.

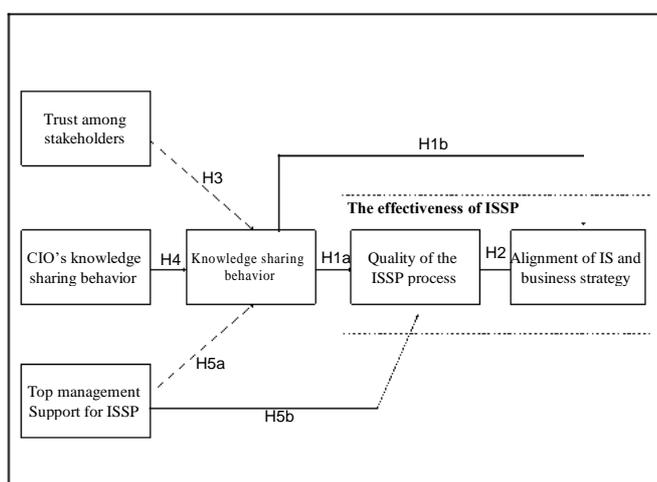


Figure 1 Research model

Each stakeholder group in the IS/IT strategic planning process is perceived to have specific backgrounds and expertise (Earl, 1993). Consequently, knowledge sharing among different groups not only generates creative ideas but also reduces conflict (Hackney and Little, 1999). Furthermore, a general consensus and acceptable results for all group members can be reached, ensuring that all members of an organization participate effectively in IS/IT strategic planning. Moreover, effective knowledge sharing has been presented in the literature as one means for improving innovations and performance. Connelly and Kelloway (2003) argued that knowledge sharing is emerging as an important concept and is often cited as an antecedent of innovation. Huysman *et al.* (1994) indicated that ISSP is an innovative process. Knowledge sharing behavior among interested groups could help an organization to enhance its managerial capability, change organizational structures and reach a consensus of IS use and management (Kearns and Lederer, 2000). It is generally argued that these knowledge sharing influence creative IS/IT planning in organizations (Hackney and Little,

1999; Salmela *et al.*, 2000). Significantly, a successful IS/IT strategic planning process should consider knowledge sharing behavior among different stakeholder groups. Thus, we hypothesized:

H1a: Knowledge sharing behavior significantly and positively affects the quality of the ISSP process.

H1b: Knowledge sharing behavior significantly and positively affects the alignment of IS and business strategy.

As technology continues to evolve, grow and become increasingly complex, the information systems strategic planning process becomes complex and difficult to handle. Alignment of the information systems strategy and business strategy is generally considered the key to successful information systems strategic planning (Henderson and Venkatraman, 1993; Earl, 1993; Segars and Grover, 1999). Notably, a comprehensive information systems strategic planning process produces a higher quality information system plan that ensures better plan implementation, and in turn improves the alignment of the information systems and business strategies (Lederer and Salmela, 1996; Pun and Lee, 2000). Thus, we hypothesized:

H2: The quality of the process significantly and positively affects the alignment of IS and business strategy.

“Trust” has been as a key factor for social exchange process (Blau, 1964; Munch, 1993). There is no any contract or restriction exist in knowledge sharing among stakeholders in the planning process, therefore; trust among stakeholders is key to successful knowledge sharing. The trust will be established if the benefits can be get with each other in exchange process, then it will facilitate the knowledge sharing among stakeholders. It is likely that different stakeholders will have different reasons for wanting an ISP study to take place (Galliers, 1991; Ruohonen, 1991). CIOs being mainly concerned with the implementation of strategy, while user managers are not satisfied with the ISSP process perhaps because they seek more influence and general managers emphasize SISP method issues because they find strategy-making far from easy (Earl, 1990; Ruohonen, 1991). The goals and consensus will not be achieved if the trust among stakeholder groups is not existed. Thus, we hypothesized:

H3: Trust among stakeholders significantly and positively affects knowledge sharing behavior among stakeholders.

The CIO has been described as the corporate officer who truly understands the interconnection of the information flow to the business (Benjamin *et al.*, 1985; Stephens *et al.*, 1992). In some firms, IT expertise and knowledge exist in the minds of a few senior IS managers and are not communicated to others in the

firms. Top managers or user managers do not know enough about the strategic use of IT and are not familiar with IS/IT strategy issues and should learn more about IS/IT issues. Better CIO's knowledge sharing not only facilitates the integration of different views from stakeholder groups, but also facilitate these stakeholders more cooperative. From the above, we can understand that CIO's knowledge sharing can promote organizational members to understand what IS and IT can do for the company and be willing to take advantage of IS/IT opportunities when they arise, and then further promote the intention of sharing knowledge for other stakeholders. Thus, we hypothesized:

H4: *CIO's knowledge sharing behavior significantly and positively affects knowledge sharing behavior among stakeholders in the planning process.*

Top management support is considered a prerequisite for every successful IS discipline. Beatty *et al.* (2001) indicated that as in all innovative endeavors in the organization, top management support is extremely important. Earl (1993) also indicated that top management support is necessary for a successful ISSP since top management is responsible for a wide range of organizational processes and activities. Top management must not only realize that IT is not a panacea for all organizational problems, but also should be viewed IT as a resource to be deployed judiciously to support or influence business strategies in terms of streamlining business operations, reengineering business processes, forging electronic links with suppliers and customers, etc (Teo and Ang, 1999). Since top management support of the importance of IT/IS is a prerequisite for securing the commitment and involvement of top management. Thus, we hypothesized:

H5a: *Top management support for ISSP significantly and positively affects knowledge sharing behavior among stakeholders.*

H5b: *Top management support for ISSP significantly and positively affects the quality of the ISSP process.*

4. RESEARCH METHOD

4.1. Sample and Data collection

The sample adopted was the Corporate 1000 (the 1000 largest manufacturing and service companies in Taiwan), published by Commonwealth Magazine in 2003. To ensure the questionnaire was received by IS directors and to encourage a better response rate, two research assistants spent about three weeks telephoning these 1000 companies. Only those companies with formal IS departments qualified as participants. Such firms were then asked to provide the name of the IS director to whom a questionnaire should be mailed. Using this procedure, a list of 805 firms from various industries was compiled. Questionnaires were mailed to the 805 IS

directors. A cover letter explaining the objective of the study and a stamped return envelope were enclosed. Follow-up letters were sent about three weeks after the initial mailings. The decision to use the IS executives as informants herein is supported by previous research conducted by Segars and Grover (1998) and Gottschalk (1999).

4.2. Measure development

The research variables were defined as briefly as possible with multiple indicator items. From the literature on social exchange, knowledge management and IS management theory, we adopted the variables that have been used and validated by other researchers. All variables were measured with multiple items on a five point Likert-type scale, ranging from (5) strongly agree to (1) strongly disagree.

4.3. Pre-testing

The questionnaire was refined through two rounds of rigorous pre-testing. The pre-testing process focused on instrument clarity, question wording and validity. During the first round of pre-testing, five MIS doctoral students and three MIS professors were interviewed. During the second round of pre-testing, a revised questionnaire was pre-tested by three senior IS executives from manufacturing, banking and retail industries.

5. RESULTS

5.1. Sample characteristics

Of the 805 questionnaires distributed, 151 completed usable questionnaires were returned, for a response rate of 19%. The respondents are all information systems executives, and had worked in the information systems field for an average of 15.3 years. The respondents came from diverse industries, with manufacturing representing 17.1%, banking/finance/insurance 25.8%, computers/communication 23.8%, and the remainder coming from various other backgrounds such as real estate, construction, health and transportation. This result implies that ISSP is carried out in a wide variety firms.

5.2. Reliability and validity of research variables

Internal consistency (Cronbach's alpha) was calculated in order to assess the reliability of all constructs. The results in our study indicate that all the constructs are greater than 0.7. The constructs are therefore considered to exhibit adequate reliability (Nunnally, 1978). The content validity of questionnaire was established through a series of personal interviews with multiple IS executives. Construct validity was determined using factor analysis of the items comprising each construct. Principal component factor analysis with VARIMAX (orthogonal) rotation was used to determine if all

items measuring a construct cluster together and selection of factors with eigenvalues greater than one. Items with loadings of less than 0.5 on any factor were dropped from subsequent analyses. The results of the factor analyses for independent variables confirm that each construct is distinct from other constructs.

5.3. Hypothesis testing

The hypothesized relationships depicted in research model were tested using multiple regression analysis. No apparent colinearity problem among independent variables exists, as the variance inflation factors (VIF) for all independent variables are smaller than 3 (Neter *et al.*, 1997).

As predicted by H1a, knowledge sharing behavior significantly and positively affects the quality of the ISSP process ($\beta=0.22$, $p<0.01$). Moreover, knowledge sharing behavior significantly affects the alignment of IS and business strategy ($\beta=0.29$, $p<0.001$), supporting H1b. Furthermore, the quality of the ISSP process has a significant positive effect on the alignment of IS and business strategy. Thus, H2 is supported. As proposed H3, trust among stakeholders shows a positive relationship with knowledge sharing behavior ($\beta=0.31$, $p<0.001$). Consequently, H3 is supported. Moreover, H4 is supported since the CIO's knowledge sharing behavior has a significant positive effect on knowledge sharing behavior ($\beta=0.17$, $p<0.05$). Finally, top management support for ISSP has a strong significant effect on knowledge sharing behavior and the quality of the ISSP process, so H5a ($\beta=0.43$, $p<0.001$) and H5b ($\beta=0.37$, $p<0.01$) are supported.

6. DISCUSSION

Results of this study demonstrate that knowledge sharing behavior significantly affects the quality of the ISSP process (H1a) as well as alignment of IS and business strategy (H1b), implying that as ISSP become more important to organizations, effective knowledge sharing among stakeholders is required. ISSP requires views from a range of stakeholders and the knowledge sharing of these stakeholders to achieve common goals. Furthermore, different managerial groups should participate in planning, increasing mutual understanding and sharing the use of the information (Lederer and Mendelow, 1988; Reponen, 1993). ISSP is an iterative, ongoing and complex process that cannot be handled by just one person in an organization (Auer and Reponen, 1997). An effective knowledge sharing mechanisms can help planning participants share their implicit and tacit knowledge, and to achieve their decision-making goals. According to a recent study based on resource-based theory conducted by Kearns and Lederer (2003), the results showed that knowledge sharing processes of strategic IT alignment influence two sets of outcomes (including business plan reflecting the IT plan as well as

IT plan reflecting the business plan) and then yield competitive advantage for organizations. The findings of this study suggest that both practitioners and researchers should direct significant effort understanding shared knowledge, the factor which had the strongest influence on the quality of the ISSP process and alignment of business and IS strategies.

Trust among stakeholders was originally hypothesized to be related to knowledge sharing behavior. The results provide substantial evidence of this hypothesis, implying that mutual trust among stakeholders is essential to successful teamwork. Trust is a major component of the cooperative competency that is established in organizational theory (Sivadas and Dwyer, 2000). It has defined as the confidence a department (or stakeholder) has in the ability and motivation of the other department (or other stakeholders) to produce positive outcome for the organizations. The key objective of ISSP is to establish a strategic IS plan that is satisfactory to different groups. However, ISSP is so complex that it cannot be accomplished if the organization lacks mutual-trust among stakeholders. During planning, stakeholder groups are mutually interdependent and each has specific tasks and responsibilities. Accordingly, organizations require mutual-trust to achieve planning goals and ensure that the IS plan is implemented as expected.

CIO's knowledge sharing behavior was found to be positively and significantly associated with the knowledge sharing behavior among stakeholders (H4). Similar findings have proposed by previous researchers such as Stephens *et al.* (1992) suggest that the CIO operates as an executive rather than a functional manager. They further indicated that CIO is an active participant in strategy planning and acts as a bridge between the information group, the functional areas, and external entities. According to Teo and Ang (2001), senior information systems executives fail to win top management support because they talk in technical jargon that top management do not understand. IS executives should bear in mind that top management are more interested in knowing how IT can help leverage the company's competitive position. As a result, we can understand that CIO may focus on strategic issues rather than technical issues while sharing his IT knowledge with CEO.

Top management support for ISSP was found to positively and significantly affect the knowledge sharing among stakeholders and the quality of the ISSP process. The result was consistent with early findings of similar studies that indicated that top management is found to be important facilitator for effective IS strategic planning (Premkumar and King, 1992).

The results indicate that the casual relationship exist between the quality of the ISSP process and alignment of IS and business strategies. That is, the alignment

between IS strategy and business strategy improves with the quality of the ISSP process. The results are consistent with reports in the literatures (King, 1988; Premkumar and King, 1994; Lederer and Salmela, 1996). Since the quality of the ISSP process is a prerequisite of planning alignment, the factors affecting the quality of the ISSP process must be understood when undertaking ISSP. This study hypothesized the knowledge sharing behavior positively and significantly affects the quality of the ISSP process, the empirical evidence support this hypothesis.

7. CONCLUSIONS

To achieve the consistently planning objectives, knowledge sharing is necessary to IS/IT strategic planning. One important question is that of how knowledge sharing can be achieved in the IS/IT strategic planning process. According to Nambisan *et al.* (1999), numerous organizational mechanisms exist that can enhance knowledge sharing and transfer, including IT steering committee and strategic IS/IT team. Based on Mentzas (1997), this study suggested that four different types of strategic IS/IT teams should be specified when implementing an IS strategy, namely, functional, technical, guidance and coordination teams. Additionally, information technology can play a central role in knowledge sharing process (Hislop, 2002). In the ISSP process, some IT applications may help planning participants to share and transfer their knowledge. Such applications may include groupware, intranet and web-based applications (Bai and Lee, 2003). Furthermore, the cross-participation (CEO participate in IT planning and CIO participate in business planning) is necessary to elucidate the tacit knowledge that often remains undiscovered and is not shared in the organizational knowledge base (Johannessen *et al.*, 2001) and to make this personal knowledge explicit at the organizational level (Kearns and Lederer, 2003).

This study has the following limitations: First, the subjects of this study are IS executives in Taiwan. Consequently, cultural differences may exist between Taiwan's situation and other countries. Second, this study use CIO as an informant. The CEO was not used as an informant in this study **because** of the CEO possibly not being familiar with some of the detailed characteristics of information systems strategic planning examined in our research variables. The effectiveness of ISSP in this study includes technology-led measure items, thus making it difficult for CEOs to respond knowledgeably. Previously, most respondents in ISSP empirical research were CIOs (Premkumar and King, 1994; Segars and Grover, 1999; Gottschalk, 1999). In contrast, IS executives are more likely to be aware of and knowledgeable about the research variables, especially the measurement of ISSP success. Third, the sampling population included large businesses in Taiwan. However, the IS maturity of large organizations is greater than that of SMEs, meaning that the results herein may have

limited generalizability for SMEs.

Future research should focus on four areas, as motivated by limitations of the present study. First, future researchers might attempt to understand the conclusions about the knowledge sharing factors of this research through structured interviews in case studies of IS directors from ongoing or recently completed ISSP projects. Researchers might ask subjects why these factors are associated with the effectiveness of ISSP. Second, future researchers could consider more general factors that affecting knowledge sharing behavior such as task coordination, rewards and political behavior. Third, the subjects of this study are IS executives in Taiwan. Cultural differences may exist between Taiwan and other countries. Future research should be similarly carried out in other countries.

ACKNOWLEDGMENTS

The authors would like to thank the National Science Council of Taiwan, China for financially supporting this research under Contract No. NSC92-2416-H-147-002.

REFERENCES

- [1] Auer, T. and Reponen, T. (1997) "Information systems strategy formulation embedded into a continuous organizational learning process", *Information Resources Management Journal*, 10(2), pp.32-43.
- [2] Beatty, R.C., Shim, J.P. and Jones, M.C. (2001) "Factors influencing corporate web site adoption: a time-based assessment", *Information and Management*, 38, pp. 337-354.
- [3] Benjamin, R.I., Dickinson, C.J. and Rockart, J.F. (1985) "Changing role of the corporate information systems officer", *MIS Quarterly*, 8(3), pp. 177-188.
- [4] Blau, P. M. (1964), *Exchange and power in social life*, New York: John Wiley and Sons.
- [5] Connelly, C. E. and Kelloway, E. K. (2003), "Predictors of employees' perceptions of knowledge sharing culture", *Leadership and Organizational Development Journal*, 24(5), pp. 294-301.
- [6] Cunningham, N. (2001) "A model for the role of information systems in strategic change within healthcare organizations", *Organizational Development Journal*, 19, Spring, pp.93-105.
- [7] Earl, M.J. (1989) *Management Strategies for Information Technology*, Prentice Hall, Englewood Cliffs, New York.
- [8] Earl, M.J. (1990) "Approaches to strategic information systems planning: experience in twenty-one United Kingdom companies", *Proceedings of 11th International Conference on Information Systems, Copenhagen*, Denmark, pp. 271-277.
- [9] Earl, M.J. (1993) "Experiences in information systems strategic planning", *MIS Quarterly*, 17(1),

- pp.1-24.
- [10] Galliers, R.D. (1991) "Strategic information systems: myths, reality and guidelines for successful implementation", *European Journal of Information Systems*, 1(1), pp. 55-64.
- [11] Gottschalk, P. (1999) "Implementation predictors of strategic information systems plans", *Information and Management*, 36, pp.77-91.
- [12] Hackney, R. and Little, S. (1999) "Opportunistic strategy formulation for IS/IT planning", *European Journal of Information Systems*, 8, pp.119-126.
- [13] Henderson, J.C. and Venkatraman, N. (1993) "Strategic alignment: leveraging information technology for transforming organizations", *IBM Systems Journal*, 32, pp.4-16.
- [14] Hislop, D. (2002) "Mission impossible? Communicating and sharing knowledge via information technology", *Journal of Information Technology*, 17, pp. 165-177.
- [15] Huysman, M.H, Fischer, S.J and Heng, M.S. (1994) "An organizational learning perspective on information systems planning", *Journal of Strategic Information Systems*, 3, pp.165-177.
- [16] Johannessen, J. Olaisen, J. and Olsen, B. (2001) "Mismanagement of tacit knowledge: the important of tacit knowledge, the danger of information technology, and what to do about it", *International Journal of Information Management*, 21(1), pp. 3-20.
- [17] Kearns and Lederer (2003) "A resourced-based view of strategic IT alignment: how knowledge sharing creates competitive advantage", *Decision Sciences*, 34(1), pp. 1-29.
- [18] Lederer, A.L., Salmela, H. (1996) "Toward a theory of strategic information systems planning", *Journal of Strategic Information Systems*, 5, pp.237-253.
- [19] Lederer, A.L. and Sethi, V. (1996) "Key prescriptions for strategic information systems planning", *Journal of Management Information Systems*, 13(1), pp.35-62.
- [20] Lee, G.G. and Pai, R.J. (2003) "Effects of organizational context and inter-group behaviour on the success of strategic information systems planning: an empirical study", *Behaviour and Information Technology*, 22(4), pp. 263-280.
- [21] Levy, M., Powell, P. and Galliers, R. (1999) "Assessing information systems strategy development frameworks in SMEs", *Information and Management*, 36, pp.247-261.
- [22] Mentzas, G. (1997) "Implementing an IS strategy - a team approach", *Long Range Planning*, 30(1), pp.84-95.
- [23] Munch, R. (1993), *Sociological theory: from the 1850s to the present*, Chicago: Nelson-Hall Publishers.
- [24] Nambisan, S., Agarwal, R. and Tanniru, M. (1999) "Organizational mechanisms for enhancing user innovation in information technology", *MIS Quarterly*, 23 (3), pp.365-395.
- [25] Neter, L., Wasserman, W. and Kutner, M.H. (1997) *A linear model*, 3rd Edition, Richard D. Irwin, Homewood.
- [26] Nonaka, I. (1994) "A dynamic theory of organizational knowledge creation", *Organizational Science*, 5(1), pp.14-37.
- [27] Nunnally, J.C. (1978) *Psychometric theory*, McGraw-Hill, New York.
- [28] Premkumar G. and King, W.R. (1992) "An empirical assessment of information systems planning and the role of information systems in organizations", *Journal of Management Information Systems*, 9(2), pp.99-125.
- [29] Premkumar G. and King, W.R. (1994) "Organizational characteristics and information systems planning: an empirical study", *Information Systems Research*, 5(2), pp.75-109.
- [30] Pun, K.F. and Lee, M.K.O. (2000) "A proposed management model for the development of strategic information systems", *International Journal of Technology Management*, 20(3/4), pp. 304-325.
- [31] Reponen, T. (1993) "Strategic information systems--a conceptual analysis", *Journal of Strategic Information Systems*, 1(2), June, pp.100-104.
- [32] Ruohonen, M. (1991) "Stakeholders of strategic information systems planning: theoretical concepts and empirical examples", *Journal of Strategic Information Systems*, 1(1), pp. 15-28.
- [33] Sabherwal, R. (1999) "The relationship between information system planning sophistication and information system success: an empirical assessment", *Decision Sciences*, 30(1), pp.137-167.
- [34] Salmela, H., Lederer, A.L. and Repoen, T. (2000) "Information systems planning in a turbulent environment", *European Journal of Information Systems*, 9(1), pp.3-15.
- [35] Segars, A.H. and Grover, V. (1998) "Strategic information systems planning success: an investigation of the construct and its measurement", *MIS Quarterly*, June, pp.139-163.
- [36] Stephens, C.S., Ledbetter, W.N., Mitra, A. and Ford, F.N. (1992) "Executive or functional manager? The nature of the CIO's job", *MIS Quarterly*, December, pp. 449-467.
- [37] Teo, T.S.H. and Ang, J.S.K. (2001) "An examination of major IS planning problems", *International Journal of Information Management*, 21, pp. 457-470.
- [38] Teo, T.S.H. and King, W.R. (1997) "Integration between business planning and information systems planning: an evolutionary-contingency perspective", *Journal of Management Information Systems*, 14(1), pp.185-224

The Roots of the Uncertainty in the Enterprise Tech-innovation Process under the Net-Environment

Xinghua Dang, Junfeng Jiang

School of Business Administration, Xi'an University of Technology, Xi'an 710048, China
gsg1001@xaut.edu.cn; nelsonkiang@yahoo.com.cn

ABSTRACT

This paper applies evolutionary view to investigate the enterprise tech-innovation process and study the uncertainty of the routine, net environment and the innovational technique. In the end the measurement of risk is described as the uncertainty characteristic then induced the coherent characteristic between the risk and the uncertainty, which can be applied in the theory on the innovation management.

Keywords : information flow; uncertainty; net-environment

1. INTRODUCTION

It is not an isolated event that enterprise's technology is achieved newly under the network environment. It exists as the result of the course. During the process of its organization realizing, uncertainty is the most essential characteristic of technological innovation. This text believe that its ultimate behavior can be reflected by the deviation of the achieved degree of the strategic destination at the given time point.

Among the scholars, there are lots of viewpoints that connect each other and set against to the uncertainty roots in the tech-innovation. Theory of Impacting perspective [Markus, L. & Robey, 1988; Thompson, 1967]^{[1] [2]} believed that the quantity of technological information and the technological means constitute the uncertain source; And believed that, the exquisiteness not only influences the professional technological innovation course, but could promote the organization ability of the technological innovation during the course of enterprises tech-innovation with the formers' promotion. Theory of selective perspective and theory of conflict [Orlikowski and Yates, 1994]^[3] believed that technology is the product of human's interaction, design and allocation. All the communication techniques which applied by the organization results from the complex social interaction, among which the main conflict s come from the individual and organizational framework^[4].

In any case, this paper thinks the technological innovation as a coordination and uncertain course. We regard the incomplete knowledge and bounded rationality as the characteristic in the technological innovation course. Especially, because the object that tackles is incomplete restructure of the existing information, knowledge, technology, etc., during which the new knowledge must be created. Many domestic and foreign scholars have strong interests in the risk measurement of the enterprise's innovation course. Among the numerous viewpoints, a agreement is

achieved that risk founds on the goal-structure, which an aprioristic measurement index system is demanded while the index system couldn't be constructed by the economists and management experts at present^{[6] [7] [8]}.

2. THE ROOTS OF THE UNCERTAINTY IN THE ENTERPRISE TECH-INNOVATION PROCESS UNDER THE NET ENVIRONMENT

The concept of the uncertainty in the enterprise tech-innovation process refers to the total deviation of the current conditions with included in each of the decision points of the enterprise tech-innovation process and the choices with difference effected on enterprise by the very points. The result it directly caused is the total difference of the tech-innovation activity and the final decision's difference (relative).

Any course of the enterprise tech-innovation under the net-environment has involved three most essential objects: enterprise's structure, net-environment and innovation technology as for an object. The tech-innovation process is not only the development of the technology but also the changes of the social productivities' structure.

2.1 The structure of enterprise

The structure of enterprise is an economy system with a unique becoming history. While the enterprise engaged in tech-innovation activity, it must have the resources of two respects at least: 1, the information resources precipitated in past history; and 2, the convention that forms by searching in its history. The information resources of enterprises have the characteristics of two respects: 1, the size of its information amount; 2, the quality that information has (accuracy). According to information theory, the amount of information could be described according to its entropy, and according to Qiu Yuanhua's viewpoint: the quality that information has could be described by its complex entropy. In a word, the amount of information resource and its accuracies at

the time t could be described by a complex entropy $H(x_t)$ ^[9].

$$\begin{aligned}
 \text{Definition: } H(x_t) &= \sum_{k=1}^n |P(x_t) \ln P(x_t)| \\
 &= \sum_k |P(x_t) [\ln P(x_t) + i \arg \ln P(x_t)]|, \quad (2-1)
 \end{aligned}$$

In which: $x \in C$. So $H(x_t)$ indicates the above mentioned two respects of the information. It postulates the vector of information input in the course of the enterprise tech-innovation is $X_t^{n_1} = (x_{t1}, x_{t2}, \dots, x_{tn_1})$

The output of a given information vector transacted by the enterprise's convention is:

$X_o^{n_2} = (x_{o1}, x_{o2}, \dots, x_{on_2})$. The transformation for a

given input information vector in the course of the enterprise tech-innovation is $W_t = \bigcup_{k=1}^{n_3} w_{tk}$, which

means that at the time t , the transformation which is suitable to any input information is selected in the

ensemble: $W_t = \bigcup_{k=1}^{n_3} w_{tk}$. The probability for a given

w_{tk} to transform the information is P_{tk} . Under the above-mentioned condition,

the equation: $X_o^{n_2} = X_{it}^{n_1} W_{tk}$ exists, in which

$k \in (1, 2, \dots, n_3)$, i.e., when the input information vector

is $x_{it}^{n_1}$, enterprise chooses w_{tk} in $W_t = \bigcup_{k=1}^{n_3} w_{tk}$ to

transform, the information's output is $X_o^{n_2}$; At the same time, from the viewpoint of enterprise's evolution,

this kinds of enterprise could choose the uncertainty

of w_{tk} in $W_t = \bigcup_{k=1}^{n_3} w_{tk}$ to make the enterprise

tech-innovation process could be indicated by an iterated function system with probability, that is

$IFSP_E : (R^{n_2}, w_{t1}, w_{t2}, \dots, w_{tm_3}; P_{t1}, P_{t2}, \dots, P_{tm_3})$ while.

$$\sum_{k=1}^{n_3} P_{tk} = 1. \quad (2-2)$$

The above-mentioned $IFSP$ is not only a fractal, but also because of its satisfactory characteristic of self-resemblance, it could explain the complexity in the evolution of the enterprise's becoming history very well ^[10]. Now discuss the becoming evolution of the convention in the course of enterprise tech-innovation.

The transform set is one cluster of enterprise's routine that has acted on outside and inside existing information

resources $W_t = \bigcup_{k=1}^{n_3} w_{tk}$. It decides the information's

characteristic and the evaluative course of itself at the

next time point. This paper believes that the evolution of enterprise's routine is not depended on the result that enterprises calculate in advance, but limited by the result that chosen by society network. Therefore, once enterprises choose w_{tk} to transform, when w_{tk} is used on next step, it must has already absorbed the result this time. Embodying in mathematics, it

$$w_{(t+1)k} = w_{tk} + \Delta w_{tk} \quad (2-3)$$

In which t_0 is the cycle of choosing w_{tk} ; Δw_{tk} is the emendation matrix, which reflects the revision of routine because of choosing w_{tk} this time, namely the routine which the change causes is revised.

2.2 The network environment

The network environment represented by Internet dwindles the time lag greatly. Considering the network environment as a whole behavior characteristic for the system, the paper believes that the network environment could be regarded as an independent behavior subject because of the network imbedding formed in the course of the building of the innovation's object. Its behavior characteristic is similar with the enterprise's structure: namely it has certain information and convention. At the same time information is dealt with by the routine, the routine is evolving in this course. Besides of the fractal structure which is similar with the enterprise's structure, the pre-requisite factor of network environment is to choose function. This requires looking for a parameter of scaling the characteristic to confirm the choosing function.

Firstly, $IFSP$ structure is used to indicate the self-evolution course of the network, $IFSP_N : (R^{m_2}, w_{t1}, w_{t2}, \dots, w_{tm_3}; P_{t1}, P_{t2}, \dots, P_{tm_3})$ while,

$$\sum_{k=1}^{m_3} P_{tk} = 1 \quad (2-4)$$

On the other hand, in order to indicate the heredity and variation of the convention of the network, similarly

with enterprise's structure $w_{(t+1)k} = w_{Ntk} + \Delta w_{Ntk}$,

T_0 is the circulation cycle of w_{Ntk} chosen for the

network; Δw_{Ntk} is the emendation matrix, which reflected the revision of convention because of choosing this time, namely the convention which the change causes is revised. Finally, to resolve the choosing function and scale the competition relation between

convention and information, this paper sets about from the outside, confirms that $IFSP_N$ is chosen to be

confirmed by the radius Z of the table in the technological convention that t may include

$$\text{constantly, } Z_{(t)} = |\ln C_{tm} - \ln C_{t1}| \quad (2-5)$$

Then $Z_{(t)}$ has determined field of the choices of the information resources and convention resource in the network. where C_m is the cost of the technological information and convention S_m , C_{t1} is the cost of technological information convention S_{t1} . Because the evolution of the network makes technological diffusion faster, enterprise's competition is aggravated in the network, then this paper supposes that $Z_{(t)} \leq 0$. That means that with the evolution and development of the network, the difference on the cost of the information suitable for choosing and the routine will be smaller and smaller in the network. On the other hand, because of the variety of the network, the division among specialized departments of the technological innovation course will be more and more meticulous.

2.3 Tech-innovation as the object of development

Considering the object of enterprise's development and technological source, it is possible to have two kinds: 1).this tech-innovation is not existing in the environment of the network, namely the network environment does not have the same information as this, the existence of tech-innovation in the network environment gone on and existed by this enterprise, namely the original innovation in a normal sense; 2).this tech-innovation exists in the environment of the network, and does not exist in this specific enterprise, namely the imitation innovation in a normal sense.

Setting up the permutation of all kinds of technology when moment t of the environment of the whole network as follows: $T_{Nt} = \{S_{t1}, S_{t2}, KS_m\}$ (2-6).

This permutation Satisfies the above-mentioned assumption: namely by virtue of competition of the society, the possible technology choosing included in T_{Nt} is confirmed by the radius of the table

$$Z_{(t)} = |\ln C_m - \ln C_{t1}|, \text{ and } T_{Nt} = \{S_{it} \in T_{Nt}\}.$$

And two kinds of situations which are perhaps in of this tech-innovation of enterprise while discussing are expressed in order to make two types:

1), $S_{t+T_E} \in T_{Nt}$, namely this tech-innovation is imitation innovation. And

2), $S_{t+T_E} \notin T_{Nt}$; $S_{t+T_E} \in T_{N(t+T_N)}$, namely this tech-innovation innovates originally.

3. ENTERPRISE TECH-INNOVATION RISK MODEL

The uncertainty in the course of enterprise tech-innovation of course could be divided into two levels according to aspect: 1.the deviation of the goal object and the routine it selected to use in the structure

of enterprise tech-innovation; 2. the risk that the uncertain factor in the course.

3.1 The risk resulting from the routine

According to the above mentioned: The route where enterprise evolves can be determined $IFSP_E$, its routine is $[R^{n2} : W_1, W_2, \dots, W_{n2}]$, it is a IFS_E with a compression factor $0 \leq s < 1$, and in which IFS_E

the transformation is $w : h(x) \rightarrow h(x)$ (3-1).

Define: $w(B) = \bigcup_{n=0}^{\infty} w_{k_t}(B) \quad \forall B \in h(x)$, (3-2)

Where w_{k_t} is the compressive transform with a factor s in the space $(h(x), h(x))$ which satisfied the following:

$$h(w(B), w(C)) \leq sh(B, C), \quad \forall B, C \in h(x) \quad (3-3)$$

In which its only fixed point A satisfies

$$A = w(A) = \bigcup_{n=0}^{\infty} w_{k_t}(A), \quad \text{and } A = \lim_{n \rightarrow \infty} w^n(B)$$

$\forall B \in h(x)$. It is a fractal. Now considering evolution of enterprise, definition of the corresponding displacement transformation $S_t: A \rightarrow A$ of

$$w(B) = \bigcup_{n=0}^{\infty} w_{k_t}(B) \quad S(a) = w_{k_t}^{-1}(a), \quad \alpha \in w(A)$$

4). Dynamical system $(A; S)$ is the displacement dynamical system going with IFS_E . It can be written as following: $S(a_n) = w_{nk_t}^{-1}(a_n), \alpha \in w_n(A)$ (3-5).

Because of the sensitiveness $S(a)$, the difference of the tracks of $w(B) = \bigcup_{n=0}^N w_{k_t}(B)$ will extend with the effect of dynamical system, i.e. $h_t(w(B), w(C))$ will change. In which h is the Hausdorff distance.

Therefore, it could be measured of the possibility of the difference between the direction which indicates the enterprise innovation and the direction of network evolution in the track space $w_{(w(B), w(C))_t}$ thus leads

out conclusion 1): any uncertainty of enterprise tech-innovation is closely linked with the time t in the evolution track.

3.2 Risk caused by the uncertainty factor's

Constructed being aforesaid:

$$IFSP_E : (R^{n2}, w_{t1}, w_{t2}, \dots, w_{m3}; P_{t1}, P_{t2}, \dots, P_{m3})$$

according to the random density of model, ISFP can forms finite Markov chain. Its probability shows the probability factor in the random network chart, the

picture which outputs the result can be expressed with the matrix of making up.

We need to calculate the uncertainty in the course of the tech-innovation

$H_{I\Delta}$ shows the information flows impart entropy and the measurement of accuracy after the transformation, in which W_{kt} is the complex variable matrix.

In sum, the uncertainty of enterprise tech-innovation course can be shown by above-mentioned variables: $Q = (h_i, H_{I\Delta})$. The first part h_i has expressed the uncertainty's measurement of the object chosen in the tech-innovation; The second part $H_{I\Delta}$ is the uncertainty of the tech-innovation course while organizing.

4. THE MODEL DISCUSSING AND THE MEASUREMENT OF THE RISK

In the mathematics model of section 3, the economic meaning which represented by w_{ik}

in the model $W_i = \bigcup_{k=1}^{n_3} w_{ik}$. Now we assume that w_{ik} presents the concept of economic cost, and indicates the difference measurement of the cost of individual enterprise innovation and that of network innovation in the realistic economic life; w_{ik} provides the matrix of information shift and processing, then h_i shows the difference of information output's availability in the social network.

The uncertainty at the second level could be confirmed by using two indexes of the complex variable. 1). Information transfer; 2).Accuracy of conducting of information. Thus forms a complex variable P_{kt} is the probability that transformation matrix appears, if this information flow is expressed by one chart, its mathematics expression formula must be an adjoining matrix of an unconnected chart, which expressed once enterprise chooses some to deal with the way, it has no other things to substitute choosing in the future, namely the history determines their evolution possibilities, the border matrix of unconnected chart must could be simplified to a Jordan matrix, and restrained the appearance probability of Jordan matrix own probability so can simplify its calculation.

It must be pointed out that risk is aiming at a given goal collection. Therefore, the risk measurement could be come out throughout comparing the uncertainty's index and the given goal collection index.

Definition: The risk index of the innovation activity way

$$j \text{ is: } R_j = 1 - \frac{\sum_{i=1}^{h-1} Q_{ij}}{Q_0(h-1)}, \tag{4-1}$$

In which: Q_0 is the given object restriction collection; h is the Q_0 's mark of the position according to Q_0 . The choice could be decided by the following expressions:

As for the given project j , the position form risk is $R_x^{(1)} = r_x^{(1)} r_x^{(2)}$, (4-2)

Risk of break even

$$\text{is: } R_x^{(2)} = P(w \subset t) = \int_{-\infty}^t f(w) d(w) = F(t) \tag{4-3}$$

Risk of aggressive is:

$$R_x^{(3)} = P(w \subset T) = \int_{-\infty}^{+\infty} f(w) d(w) = F(T) \tag{4-4}$$

$$\text{In which } r_x^{(1)} = - \int_{-\infty}^{+\infty} \ln f(w) dF(w), \tag{4-5}$$

$$r_x^{(2)} = \begin{cases} 0 & T < \bar{x} - k\sigma_x \\ \frac{T + k\sigma_x - \bar{x}}{T - t + 2k\sigma_x} & T \geq \bar{x} - k\sigma_x \end{cases}, \tag{4-6}$$

$f(w)$ is the density distribute function for the restraining goal from collecting the element, its normalized deviation, \bar{x} is the mean value. Its density distributing function is lead out with corresponding density evolution method by above-mentioned *IFSP*.

5. CONCLUSION

Throughout the research to the roots of the uncertainty and measurement of the risk in the enterprise tech-innovation process under the net environment, the paper's conclusions are as followed:

- 1).The uncertainty factor in the enterprise tech-innovation is decided by enterprise routine: the network and the innovation object ;
- 2).Because enterprise is a organic structure ,the using of fractal system to describe its evolution and the characteristic of self-similarity is felicitous ;the measurement of the tech-innovation risk is come out from the uncertainty of the enterprise tech-innovation and the given object restriction collection ;
- 3).Because the universal mathematics tools fail to appear, different mathematics tools are needed in the course of analyzing and researching enterprise tech-innovation process. A large amount of research on constructing and evolving the method and using with the density in *IFSP* needed, it is a focal point of next research to find out its measurable triggers value.

ACKNOWLEDGEMENT

ICEB2004 is supported by the National Natural Science Foundation of China.

REFERENCE

- [1] Markus, & Robey, . Information Technology and Organizational Change: Causal Structure in Theory and Research, *Management Science*, pp. 583- 598. 1988
- [2] J. D Thompson: *Organizations in Action*, McGraw-Hill, 1967
- [3] Wanda J. Orlikowski and JoAnne Yates, "Genre Repertoire: Examining the Structuring of Communicative Practices in Organizations," *Administrative Science Quarterly*, 39 (1994), 541-574
- [4] Bettina S.T. B chel: *Greating Knowledge Organization* Palgrave.2002.
- [5] R. Williams and D. Edge: *The social shaping of technology* Edward Elgar.2000
- [6] Jan Cobbenhagen Cheltenham: *Successful Innovation* Northampton Ma. USA 1999
- [7] David J. Teece: *Profiting from the Technological Innovation* Research Policy 15, Elgeriel Science.1986
- [8] R.S. Rosonbloom: *A Distributive Technological Competences and Capabilities* Harvard Business School Case 9-674-027,1973
- [9] Qiu Yuanhua: *Management decision and Applied entropy study* (in Chinese), Mechanical Press, 2002
- [10] Li Shuigen, Wu Jitao: *Fractal and wavelet* (in Chinese), Science Press . 2002

Transformation & Countermeasures on Business Operation Models in Uncertain Environment

Baozhen Li¹, Nana Li², Xianfeng Liu²

¹School of Public Management, Shandong University at Weihai, Weihai 264209, China
Tel: 86-631-5688630, Fax: 86-631-5688375, baozhenli@126.com

²School of Business Management, Southwest Petroleum Institute, Chengdu 610050, China
Tel: 86-28-83032180, Fax: 86-28-83032180, linanana@21cn.com

ABSTRACT

In an uncertain environment, esp. facing a market transformed from sellers' market to buyers' market, each enterprise focus not just on competitors as traditionally, but on external system sources such as customer and government and inner system sources such as employee and material etc. As the globalization-economy, the market emphasizes and the influence of a government to an enterprise changed, and the importance of knowledge management in an enterprise's inner system emerged. Under this situation, the author thinks that we should not view an enterprise as a simple machine or just an open system any more, while a complex and conformity concept should be applied to research its operation model.

The traditional business operation model (BOM) is a closed "produce - sell" model. It focuses planning and decision, which means there should be a plan first then do it. In fact, the traditional BOM is a kind of initiative-passive model, the corresponding main management theory is management process and its function. While the modern BOM is an open "perceive - response" model. It emphasizes knowledge and intelligence, which means perceived something first and then feedback. Essentially, it is a passive-initiative model and the main relevant management theory is competition intelligence and knowledge management. Based on these analysis, the author think that relative to traditional businesses try to keep stabilization during a proceeding phase, i.e. quantitative change, the modern businesses should transform the perceived changes to profitable responses, i.e. qualitative change. We can use a loop course of "induction - explanation - decision - running - again induction" to explain it. In addition, facing the complexity, the management should pay more attention to "situation and harmony", but not just to "order and control".

Keywords: Uncertain Environment, Business Operation Model (BOM), Knowledge Management

1. FIRMS IN UNCERTAIN ENVIRONMENT

Almost all enterprises have realized that coming with commercial activities are the increasing transformation and unpredictability in the information age. In uncertain social environment, the objective a firm gamed with and the focus its gamed on both have been changed. To the external environment, the market transformed from sellers' to buyers' and the main problems that marketing faced are how to do a segmentation and how to operate the business according to customers' different demands. In the point of competitive strategy, there are so many discussions in traditional strategic theories, such as a series of competition theories from Porter. However, in the dynamic information age with more and more uncertainty, competitor will become copartner and strategies of cooperation and competition are often applied alternatively and the relevant strategy theories also become focal point of concern. Government regulation is very important to an enterprise's business operation no matter in the socialism market environment like China or in the pure capitalism economic environment like the U.S.. Especially in the process of economy's globality, the government control over enterprises is much stronger but not fainter than before. To the internal environment, intangible assets are more important than tangible assets.^[1,2] Therefore, it means

that service becomes more significant than product itself from the point of output and the most important resources to an enterprise are intangible assets such as technique, trademark etc. from the aspect of input because these intangible assets' value have already exceeded that of tangible assets such as land, machine, stock-in-trade and financial assets etc. Compared with traditional economy, knowledge become the vital production factor in new economy. The status of knowledge in new economy is just like capital in industrial economy and land in agricultural economy.^[3,4,5]

Even early in 1962, Simon pointed out that: "The change of pattern is inevitably to occur only if economics involves in the uncertainty research increasingly and pays more attention to the complexity of economy decision. Accordingly, business operation model (BOM) coincided with the economic circulation rules must be changed along with the increasing uncertainty of environment.

To the internal environment, an enterprise is an open dynamic non-linear system. First of all, an enterprise is an organism system communicating material, energy and information with the external environment and it has the motivity of self-perfecting and self-evolving. Second, an

enterprise is a non-linear system. This mainly attributes to the positive and negative feedback of an enterprise:^[6] The positive feedbacks interrelated to an enterprise include economies of scale, knowledge accumulation, credit and cooperation, culture and tradition, while the negative feedbacks include the empiricism and conservatism of administrators which will increase to hurt they gumption, the aging techniques because of difficulties to retreat which will lead to reduce its competitiveness, the internal nepotism coming into being which will weaken the validity of inner management, the increasing internal organizations and management levels which will affect the enterprise's efficiency, and the increasing demands for resources which will restrain the power of enlarging reproduction. To be different from a

traditional isolated system (simple machine) or an open system only emphasizing external communication function, modern enterprise should be treated as an open dynamic non-linear system. At the same time, an enterprise should give enough attention to both sides-inside and outside- and harmonizes all kinds of resources and elements.

2. THE PATTERN CHANG OF ENTERPRISE'S BUSINESS OPERATION MODEL IN UNCERTAIN ENVIRONMENT

There is some qualitative change about an enterprise's business operation model in uncertain environment. It is showed just as following table:

	Environment Condition	Organizational Structure	Operation Model	Emphases	Operation Characteristics	Management Theory in Mainstream
Traditional Enterprise	Predictable and Stable	Pyramid Multi-layer Control Rigidly	"Produce-Sell" Closed Model	Plan and Decision	Planning first then Running, Initiative-Passive	Management Process and it's Function
Modern Enterprise	Dynamic Unpredictable Unstable	Flat Multi-department 、 Flexible Unite	"Perceive-Response" Open Model	Knowledge and Information	Perceive first then Response, Passive-Initiative	Competition Information and KM

In the traditional enterprises, all kinds of operation processes are integrated only in one high-level flow mode, which pre-determined the interaction role of the processes do to each other. So just one or two decisions might be viewed as reasonable only if they are involved in specific plans or in the permissive range of schedule. Having standard operation processes, detailed order and a rigid controlled mechanism, the traditional enterprise tries to keep stable in an action phase, and want to enhance efficiency hence increase profit through repeating the same thing and learning curve effect. This business operation model is called "produce-sell" closed model. It will not response until outside environment changes are so severe that they cannot be ignored. Since this model pays attention to planning and deciding, manifests planning first and operating followed, its essence is initiative-passive. The mainstream management theory adapted to it is management process and its function.

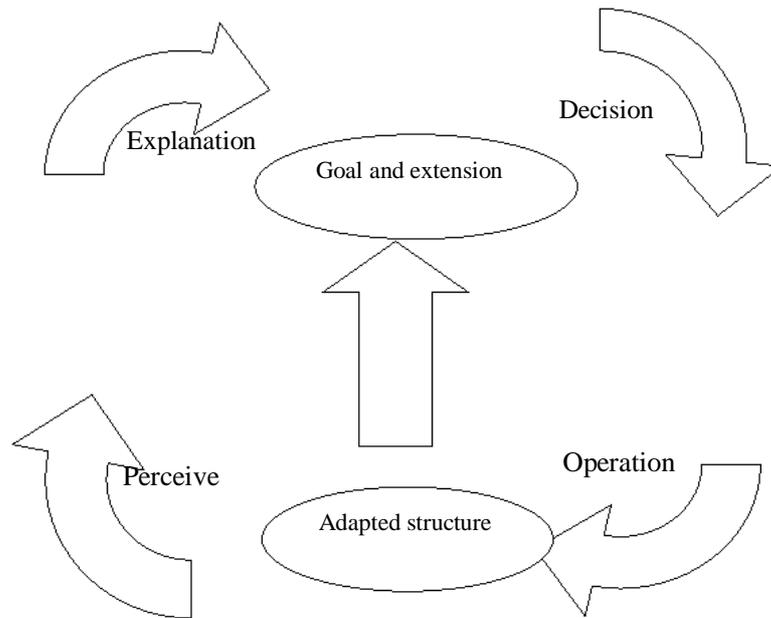
Facing complexities, modern enterprise must break the rigidity of its system and then divide the system into a few small parts, which are easier to manage. Like that, an enterprise can realize uniform action all over various functions, try to circulate in the adapted loop as quick as possible, and transform the perceived changes into a new profitable response. This business operation model (BOM) is called "perceive-response" open model, which doesn't ignore the environment's changes but focus on searching for developing signals actively, pays attention to knowledge and information, and gains great performance depending on the ability of adapting to

happence quickly. In essence, it is dynamic and unstable. So it will perceive initially and then response in action. It's a mode of passive initiative. The corresponding mainstream management theory is competitive information and knowledge management.

3. THE COUNTERMEASURE OF ENTERPRISE BUSINESS OPERATION MODEL IN UNCERTAIN ENVIRONMENT

In "sense-respond" model searching for scale adaptation, the following requirements of knowledge managements are especially important:^[7] Organization should appoint the employee who can identify effective model to the relevant posts as some elements of organization is very easy to be influenced by high-level uncertainty in environment; Providing these employees some special support information according to their posts so that their ability to identify disturb and respond rightly can be increased; It is required to described out how an organization will and how it can create value in a systemic way.

In order to realizing operation benefit, modern enterprises should continually transform the perceived changes into profitable response basing on knowledge management theory in adapted loop courses of " induction — explanation — decision — running — again induction ", that is to say, qualitative change. This course should not be like traditional businesses, which tries to keep stable in a action phase and in which only quantitative change happens. In addition, facing to the



Realizing Benefit via Knowledge Management

complexity, enterprise management should pay more attention to “environment and harmony” besides “order and control”.

REFERENCES

1. Stewart, Tom, 1997, *Intellectual Capital: The New Wealth of Organizations*. New York: Doubleday;
2. Edvinsson, Leif, and Michael S. Malone, 1997, *Intellectual Capital: Realizing Your Company's True Value by Finding Its Hidden Brainpower*. New York: Harper Business
3. Quinn, J. B., 1992, *Intelligent Enterprise*. New York: Free Press.;
4. Drucker, Peter F., 1993, *Post-Capital Society*. New York: Harper Business.;
5. Burton-Jones, A., 2000, *Knowledge Capitalism*. Oxford: Oxford University Press
6. Huaide Liu, *The Research of Uncertain economy*, Shanghai University of Finance and Economic Press, 2001
7. Stephan H. Haeckel, 1999, *Adaptive Business' Knowledge-Management*, Harvard University Press

What All Stakeholders Should Do: Creating E-Learning Communities¹

Zheng Zhao, Runtian Jing

School of Management, UESTC, Chengdu 610054, China
fanhuijiaspec@163.com

ABSTRACT

This paper examines how companies that offer e-learning products collaborate with their clients, external experts and end users. Given increased demands for more sophisticated learning products, it is becoming increasingly crucial for e-learning firms to source and exploit content, education, knowledge and expertise that are beyond the traditional boundaries of the firm. These changes raise a set of problems related to how firms can effectively interact and collaborate with other stakeholders in order to create, distribute and improve e-learning products. Based on some previous researches and the existing literature on “communities of practice”, it is proposed that “learning communities” should be established by leading firms to meet demands for new e-learning products.

Keywords: learning, e learning, learning community

This paper emphasizes on enterprises that produce digital educational and training materials, hereafter known as “e-learning” firms. E-learning is a fast growing multimedia sub-sector that has emerged to meet increased demand for digital and distance based learning and training materials in educational and workplace environments.

The focus of this paper is to research the ways in which e-learning firms collaborate and communicate with their clients, external experts and end users. These firms produce e-learning packages both for the market and for specific company clients. Given the increased demands for more sophisticated learning products, it is becoming increasingly crucial for firms to source and exploit content, education, knowledge and expertise that are beyond the traditional boundaries of the firm. This occurs in three ways. First, there is the sourcing of content from the client. Second, and increasingly, there is the need to draw advice from external learning experts. These may be experts in teaching and learning or in the subject being taught. It is now necessary for firms to open up their organization to exterior knowledge and know how, to create new collaborations. Third, given the necessity of providing effective learning that is sensitive to learner/end user needs, it is now more common to involve the end user within the development process - ensuring that e-learning products are able to engage with and enhance the learning of the end user. It involves more consultation, partnership and interaction with the end user than ever before (Russell, Calvey and Banks, 2003).

While the term “learning community” can be defined in many ways (Imel, 2001), it is used here to describe the interactions between the collection of “communities of practice” integral to the firm, and the range of external experts, clients and end users involved in the creation of

an e-learning product. Given the need for flexibility and creativity in this sub-sector (Swanson and Wise, 1997), the more firms can integrate external expertise, client creativity and learners’ knowledge and viewpoints, the more effective these “learning communities” and their e-learning products will be. However as we will see, while some successes have been realized, the strategies and pathways adopted in forming these new communities are often partial and uncertain. It’s possible to conclude that e-learning firms need to more fully engage in expanding their “learning communities” to ensure the continued production of innovative e-learning products.

1. A SOCIAL DEFINITION OF LEARNING

In a social learning system, competence is historically and socially defined. How to be a physicist or how to understand the position of the earth in the universe is something that scientific communities have established over time. Knowing, therefore, is a matter of displaying competences defined in social communities. The picture is more complex and dynamic than that, however. Our experience of life and social standards of competence of our communities are not necessarily, or even usually, congruent. We each experience knowing in our own ways. It is in this interplay that learning takes place.

Consider two extreme cases. Sometimes, we are newcomers. We join a new community. We feel like an idiot among the experts. We want to learn and apprentice ourselves. We want to become one of them. We feel an urgent need to align our experience with the competence defined. Their competence pulls our experience.

Sometimes, it is the other way round. We have been in a community for a long time. We know the rules. We are

¹ Supported by NSFC (70372032, 79900022)

thoroughly competent, in our own eyes and in the eyes of our peers. But something happens. We are sent to overseas. We go to a conference. We visit another department. We meet a “stranger” with a completely different perspective. Or we just take a long walk or engage in a deep conversation with a friend. Whatever the case may be, we have an experience that opens our eyes to a new way of looking at the world. This experience does not fully fit in the current practice of our home communities. We now see limitations we were not aware of before. We come back to our peers, try to communicate our experience, attempt to explain what we have to change how our community defines competence. We are using our experience to pull our community’s competence along.

Whether we are newcomers or experts, knowing always involves these two components: the competence that our communities have established over time, and our ongoing experience of the world as a member. Competence and experience can be in various relations to each other: from very congruent to very divergent. As two examples show, either can shape the other, although usually the process is not completely one-way. But, whenever the two are in close tension and either starts pulling the other, learning takes place. Learning so defined is interplay between social competence and personal experience. It is a dynamic, two-way relationship between people and the social learning systems in which they participate. It combines personal transformation with the evolution of social structures.

2. THE ADVANTAGES OF E-LEARNING

A considerable amount of research has been conducted on e-learning. In contrast with traditional classroom learning, e-learning brings distinct benefits to learners (Beam and Cameron, 1998; Hiltz and Wellman, 1997; McCloskey, Antonucci and Schug, 1998):

Time and location flexibility. E-learning eliminates the barriers of time and distance by offering just-in-time, on-the-job learning. It has potential to reach a global audience.

Cost and time savings. In e-learning, learners do not have to travel to a specific location. It is reported that the companies using online learning can expect an average of 50% in time savings and 40% to 60% in cost savings compared with conventional way (Khirallah, 2000).

Self-paced learning. E-learning fosters self-directed and self-paced learning by enabling learner-centric activities.

Collaborative learning environment. E-learning links each learner with physically dispersed experts and other learners, together to form an online collaborative learning community (Alavi and Leidner, 2001). By

electronic means, an e-Learning environment encourages learners to ask questions that they may not be able to ask in a conventional classroom, and to share different ideas with others more easily through online discussion forums.

Unlimited use of learning material. E-learning allows unlimited access and retrieval of electronic learning material. People can review information stored in centralized knowledge repositories over and over again. The learning material can be efficiently maintained and updated.

E-learning has become an inescapable element of business in the new economy. In 1999, companies in the United States spent \$62.5 billion on educating their employees, with more than \$3 billion spent on technology-delivered training (Khirallah, 2000). Effective and efficient learning methods are greatly required by companies to ensure employees and channel partners to be timely equipped with the latest information and advanced skills. Consequently, e-learning is being adopted by many companies to expand their learning market to previously out-of-reach employees (Wulf, 1996).

Today, thanks to the widespread access to the Internet, e-learning emerges as one of the fastest-moving trends in higher education. Thousands of online courses, including degree and certificate programs, are now being offered by educational institutions worldwide.

3. COMMUNITIES OF PRACTICE

Communities of practice are the basic building blocks of a social learning system because they are the social “container” of the competence that makes up such a system. By participating in these communities, we define with each other what constitutes competence in a given context. Communities of practice define competence by combining three elements (Wenger, 1998). First, members are bound together by their collectively developed understanding of what their community is about and they hold each other accountable to the sense of joint enterprise. To be competent is to understand the enterprise well enough to be able to contribute to it. Second, members build their community through mutual engagement. They interact with one another, establishing norms and relationships of mutuality that reflect these interactions. To be competent is to be able to engage with the community and be trusted as a partner in these interactions. Third, communities of practice have produced a shared memory of communal resources—language, routines, sensibilities, artifacts, tools, stories, styles, etc. To be competent is to have access to this resource and to be able to use it appropriately. Communities of practice grow out of a convergent interplay of competence and experience that involves mutual engagement. They offer an opportunity to negotiate competence through an

experience of direct participation. As a consequence, they remain important social units of learning even in the context of much larger systems that are constellations of interrelated communities of practice.

The firm is often seen as a key organization that can house the expertise, skills and knowledge necessary for efficient and effective e-learning production. Fransman's (1994) conception of the firm as a "processor of knowledge" (Amin, 2000) is perhaps an apposite description. E-learning firms process knowledge for their own uses but they also trade in knowledge. They draw upon the knowledge of in-house experts and, increasingly, outside subject matter experts and build this knowledge into products. They comprise sections, each with responsibility for, or claiming to own, part of the design and production process.

With these issues in mind, in recent years it has become common to refer to firms as a composite or collection of different "communities of practice". The idea of a community of practice was developed by Lave and Wenger (1990) as a theory for practice-based learning in which one could undertake "legitimate participation", to serve a kind of apprenticeship with a group of insiders in an organization; organizations being comprised of a range of different disciplinary groups or collectives, each charged with specific areas of responsibility. The theory was referred to by Brown and Duguid (1991) to support their contention that the separation of knowledge from practice is unsound. They argued that the ways in which people actually work differ from official descriptions and that "learning in working" is a better way to improve performance, and most characteristic of how firms share and develop knowledge. Wenger and Snyder (2000) later described a community of practice as a "group of people informally bound together by shared expertise and passion for a joint enterprise", with members inevitably sharing knowledge in order to solve problems in their organization. Yet, while these communities might be informal and resistant to supervision, they cannot exist without management support and structure they are bounded, to varying degrees, within the bureaucratic organization of the firm.

Given the ways in which "communities of practice" are often associated with firms working in the creative industries (Raffo et al, 2000), it is reasonable to test out how far firms were acting as "communities of practice" in the production of e-learning products. At first, some key questions need to be addressed: Can "communities of practice" operate across different organizations? Can they operate without geographical proximity? In e-learning, how does involvement with external agencies impact on the firms' "communities of practice"?

Given the increasing importance of outside experts, it became a big concern with how far agencies and forces external to the organization - whether in conflict,

cooperation or collaboration with the firm - are able to penetrate or impact upon Wenger's "locally negotiated regime of competence" and "shared histories of learning" that make up the community of practice (Amin, 2000). The e-learning firm negotiates with clients who commission products, with the external freelance experts who are selling their knowledge, the end users and with the firm's own staff who design and produce the e-learning products. The types of exchanges and relative bargaining positions of the parties to these negotiations vary according to the stage in the production. Knowledge is constantly passing through these stakeholders and the firms ought to be able to benefit from these internal and external challenges in terms of enhancing their own performance.

Further, referring to Amin's (2000) discussion of the definitions of communities of practice offered by Wenger (1998), while the key dimensions of "mutual engagement", "joint enterprise" and "shared repertoire", can be individually applied to many of the firms delivering e-learning, as well as the clients and end users—it should be discovered whether such mutuality was as pronounced within the collaborative and convergent networks and communities as necessary for the production of e-learning products. From the existing literature, communities of practice as defined is not multi or inter-organizational. The groups studied by writers on communities of practice are usually involved in discrete organizations or task based activities, for instance Wenger's group dealing with insurance claims, where decisions are negotiated within a given organizational framework but on the basis of tacit knowledge or unwritten convention. The emphasis is on close study of single organizations, rather than on the diffuse and diverse networks that make up the whole production process. In an area where management skills and knowledge for effective production can be acquired through extended and exterior communities of clients, experts and learners, where mechanisms for the creation, exchange and possession of knowledge are much harder to define and where relying on others (non-firm members) is a necessary imperative, the received notion of a community of practice begins to unravel thus it is needed to develop a more appropriate understanding of how communities of practice operate within the context of production in the e-learning.

As production of e-learning products necessarily involves members of communities working in different organizations, the firm ceases to have primacy in terms of bounding the parameters of creativity and communication, or defining e-learning products. Just as Tyre and Von Hippel (1997) refer to communal or collaborative processes and "the importance of such collaborative processes [being] that no one person embodies the requisite knowledge to comprehend complex organizational problems or the requisite variety to clarify equivocal issues", the definition of learning is no longer bound by the needs or objectives of the

individual firm. The primacy of the firm may be compromised for the greater good of the production process and the advancement of the learning community of which the firm is a member. All of these cannot be contained within the conventional understanding of the firm as a bounded set of “communities of practice” - a more open-ended conceptualization is needed in order to capture the range of communities required to produce quality e-learning materials. Extending communities of practice, into “learning communities” is one route that firms can follow to enhance the quality of e-learning products, as well as to help create a more open and reflexive attitude to learning within the firm itself.

4. DEVELOPING E-LEARNING PRODUCTS

Before examining the components of the learning community, it is useful to reflect on the particular kinds of e-learning products that companies were producing. Most commonly, firms were producing interactive CD-ROM's or Web-based products — sometimes converging the two—and while the content of products differed, three broad product types can be identified. Each involves a different level of community involvement from the four key parties involved: the firm, external experts, the client and the end user.

Organization-specific “rule based”. In certain cases, learning material is ready made and is converted into technology-based. Some of this is rule based, for example the conversion of client companies' internal regulations or induction procedures. The appreciation of how learners learn should not be underestimated, but the client companies' needs and those of their learners are relatively easy to define. Much of the work concerns creating rule sets and programming. There is some literature on instructional design (Christian-Carter, 2001). In these products the learning and learning design are largely the province of the firm and the client external experts and learners are less likely to be involved in this learning community.

Organization-specific “non-rule-based”. In the production of e-learning materials, there is a growing emphasis upon client and end user needs. Open ended and non-rule based learning is becoming more popular. Firms are therefore beginning to involve the client or end user in the creative process and this requires the involvement of new disciplines such as the learning designer in order to create or facilitate specific products and/or forms of delivery. The concentration of firms on the clients' or end users' learning requirements leads us to argue for a complex assessment of the types of creativity necessary for effective production, one that incorporates understanding of the interpretation of client needs, the design of the learning approach and the graphic design and technical elements. In these products the firm, the client and the end user are often closely involved in the design, development and evaluation -

potentially a creative learning community.

Non-organization-specific. Some e-learning products may be generic, but often broadly targeting a particular occupational, educational or industry sector. In other cases, the product is produced speculatively for general sale. Whichever the case, the company producing the products has to acquire knowledge and decide on the outcomes and objectives of the e-learning. In general sale products, the firm and its designers will usually not have any sustained engagement or full knowledge of the individual client or end users themselves. Thus, the notion that end users are engaged in the wider “learning community” is limited. In this kind of generic production, the notion of a learning community may be firm-led, drawing upon established commercial models of design expertise, learning delivery or market research.

However, it is in the case of “total-open” learning products, supported by a range of tools and e-coaches, that another type of “learning community” may come more into play. Individual or small groups of learners, from all walks of life at different times, might be invited to register or share knowledge of the product. The link between members is professional or emotional rather than organizational.

5. LEARNING COMMUNITIES

As argued, in e-learning production, the extension of a “community of practice” into a “learning community” involves integration and exchange between the firm and its internal communities with the external world - most notably exterior experts, the client and communities of end users. While client and end user might be one and the same, it is more common for companies, schools or universities to act as the client and their staff or students to be identified as the end users. By describing how each of these constituencies have a role in the production of e-learning products this paper explores how effective e-learning can be better obtained through a more open and open ended approach, one embedded within informal “learning communities”.

5.1 The firm in the learning community

While the firm, as a set of communities of practice, is engaged in formal, structured, but also informal and tacit modes of learning (Amin, 2000). The successful e-learning company is one that is pushing the learning dimension much more to the fore - both in terms of product and production process. In terms of product, the way in which this is most evident is in the increased creation of specific roles for an in-house expert in the design and delivery of learning. Often referred to as the “learning designer”, this figure acts as the designer, director and evaluator of learning needs and outcomes in the production process.

The role of a “learning designer” will vary from company to company but most crucial is an outward looking, experimental approach to learning. For the learning designer, the parameters of learning are worked out in and through the production process in a manner contingent on a range of issues including client needs, resource constraints and educational principles. It is against the background of this process of negotiation with external agencies that innovative firms have increased the development of learning dialogues and attempted to redefine and reposition the role of a “learning designer”.

Many of firms were engaging with a varied client base, producing e-learning across industrial or educational sectors. Such expansion has increased the primacy of the role of the learning designer. Good learning designers should be “experts in not knowing” and can ideally work with any type of content. Clearly a figure that can manage and manipulate any kind of content to effect and implement an effective e-learning process must be attractive in a growing e-learning marketplace.

As well as employing learning designers, in terms of production, firms are becoming more outward looking and experimental in their search to create a learning community. This involves integrating external experts into the production of e-learning products. Many firms have established their own forums and network of individual practitioners and organizations interested in e-learning.

5.2 External experts in the learning community

Given that many small firms are unable to employ a full time learning designer, and with the need for ultimate flexibility in a fast changing and uncertain market place, the role of external experts, such as freelance learning designers, evaluators and educators becomes more crucial.

The role of external experts has grown in recent years as multimedia firms with strong technical and design competence but little in the way of educational expertise are looking to integrate the “learning” dimension into the product and production process. These experts are often members of informal networks, often, though not necessarily geographically “clustered” around the commissioning firm, but bound together by a history of collaboration, shared experience and know how. Given the high levels of self-employment and freelance work in this sector, the role of the “external” expert is more salient and vital than in other, more traditional industry sectors - making them key members of any learning communities.

5.3 The clients in the learning community

The client must now play a central role in the formation and maintenance of a learning community. Attitudes

to clients varied among the firms, with some of them talking about managing their clients’ expectations and the “whole process being managing the client to accept the creative”. However, others took a different view, seeing the client’s creative contribution as a central part of developing e-learning products. Although there was general agreement that clients’ needs were often difficult to define, a problem often complicated by the clients’ lack of appreciation of the scope of e-learning - the point about collaboration is that clients become involved with the developers at an early stage in the production process and can provide creative input into the design and development of e-learning materials - while, at the same time enhancing their own understanding of e-learning within the context of their own organization. For the firms the negotiations with clients and the meeting of their needs assumed a central role.

While the extent to which clients are engaged will vary, it was clearly evident that more successful and progressive firms understood the client as central in the definition and delivery of e-learning - not merely a customer to be satisfied. This ensures that the issue of learning - for both parties - remains open and subject to creative development.

5.4 The learners in the learning community

While firms, external experts and clients can often generate productive learning communities; the involvement of end users/ learners is an area that needs more work. The needs of both individual and groups of learners are often subordinated to concerns over development and design, time-scale, budget, distribution, price and so on. But there are a number of other, more hidden reasons why end user learning needs and experiences are often secondary concerns. It may be that in the provision of tried and tested “rule based” software, learners’ needs are assumed to be simple and straightforward and unworthy of detailed consideration. Further, it is often the case that clients assume that their staff or student learning needs are homogenous. Also, firms themselves may feel uncertain in challenging client’s identification of what the learner’s needs are - particularly in the case where the client is an educational institution. Even when partnerships involve the end user, there is no guarantee that the learners’ input will be as valued as that of the firm or the experts. However, for a growing number of firms, the integration of the end user is now deemed crucial. These firms come to realize the primary focus of attention needs to be on the needs of the learners.

Not only are firms looking to expand their collaboration, an increasing tendency now aims to generate a dynamic among learners and encourage them to share learning experiences and stories, and generally to engage more widely with the providers and producers of e-learning. Creating a set of mechanisms whereby all parties can

provide feedback and engage in the production process is the ultimate goal. How this can be achieved, however, remains a crucial question.

5.5 Mechanisms for managing the learning community

How are these four disparate constituents of learning communities able to bind and act to effect the development of e-learning and e-learning products? It is indicated that certain emotional and conceptual leaps must be made to open up all stakeholders to reach an openness which requires the development of relationships with reciprocity and trust (Cross and Baird, 2000). For example, the firm needs the expert's knowledge and the expert needs to trust the firm with it. Trust and friendly collaboration are pivotal in cementing this kind of community (Ingrams and Roberts, 2000).

Geographical proximity remains important in the learning community. External experts and learners in face to face exchanges remains the best way for identifying and agreeing objectives, processes and outcomes and for helping to facilitate trust. In production, it was noted that face-to-face meetings avoided endless e-mails, course content was largely written within face-to-face meetings, and problems caused by the lack of understanding of each other's functions and of the development process itself could be overcome with face-to-face exchanges.

Where it was not possible to meet directly the clients or external experts a secondary option was to use Internet Technology to enable relationships to be maintained across distance - this was a strategy many firms employed. However, often, firms found it difficult to meet or engage with the learners or end users, face to face or at distance. But if learners are to become more central to the learning community then mechanisms for integration must be found.

Some firms now make it a priority to update knowledge, to provide toolkits for learners, interactive resource centres with archived documents, links to material and more opportunities for learners "to do things themselves". Additionally, "e-coaches" now assume a central role in cementing the dispersed body of learners into the learning community. Learners may be more engaged in feedback activities, ongoing consultation and development discussions allowing learners to create content and put it on line themselves and to generally enhance the contribution of learners in the e-learning production process.

5.6 Barriers to the formation of learning communities

After revealing some of the possibilities and strategies of new learning communities, it would be misleading to

suggest that this is a general or even a widespread process - many firms remain locked into the traditional separation of the firm from the external world, prioritise firm expertise over external knowledge and give short shrift to opening up the debate within the firm or questioning the qualities of their products. Some firms claiming that its size and attitude allowed it the luxury of debate on learning theory, was later criticized by its own staff for its concentration on deadlines and profits which discouraged staff from being innovative. It seems clear that while some learners are encouraged to use tools experimentally to construct their learning, create their own content and upload whatever they like into their shared space, the capacity of many companies to experiment, make mistakes and survive is limited.

The economic pressures placed upon a company may make it difficult to enjoy the luxury of debate on learning - this is accepted. All activities take place within budgetary, human resource and time limitations. Even if the firm wishes to develop a network of external experts, clients and learners, a decision has to be made whether to formalize and structure this network - and incur maintenance costs through communication, meetings, events and so on, or to leave it open ended and informal, but potentially losing impetus. Finally, even when learning communities are in evidence, the power relationships within that network may be asymmetrical - not everyone is able to shape the direction or definition of learning exactly as they would wish.

6. CONCLUSIONS

The creation of e-learning products involves new convergences of technology, media, skills and, increasingly, individuals and organizations themselves. The management of firms and production processes must now focus on the human aspects of these convergences - the learning community. In this article, this has led to the increased utilization of client knowledge and the involvement of the external experts, clients and end users, providing an added, challenging dimension.

As more firms enter this sub-sector and as "learning design" has the potential to become a profession in its own right, the need to acquire knowledge from outside sources will increase. This question of knowledge acquisition over distance therefore applies to firms as well as learners. There seemed to be an irony in that firms find face to face meetings essential in the development of distance and e-learning products - a necessity perhaps only partially offset by new relational communities that are emerging across geographical space. The end result is a loosely bounded learning community comprising of members from a variety of organizations and groups interacting face to face and at a distance.

As more and more firms look to enter the e-learning market, not all will come ready equipped with learning designers or a discrete learning philosophy - this may undermine the quality and effectiveness of e-learning products. But by engaging in exchange with clients, experts and learners - firms can go some way to resolve the current dilemmas of this emergent industry; namely how to provide materials that are sufficiently researched, tested and appropriate for a diverse, and fast expanding, range of end users.

REFERENCES

- [1] Alavi, M., and Leidner, D., "Technology-mediated learning: A call for greater depth and breadth of research", *Information Systems Research*, 12(1), 1-10, 2001.
- [2] Amin, A., "Organizational learning through communities of practice", *paper presented at the Millennium Schumpeter Conference*, University of Manchester, Manchester, June, 2000.
- [3] Beam, P., and Cameron, B., "But what did we learn ...?: Evaluating online learning as process", *Proceedings of the Sixteenth Annual International Conference on Computer Documentation*, Quebec, Canada, 258-264, 1998.
- [4] Brown, J.S. and Duguid, P., "Organizational knowledge and communities of practice", *Organization Science*, Vol. 2 No. 1, pp. 40-57, 1991
- [5] Christian-Carter, J., *Mastering Instructional Design in Technology-Based Training*, London: Chartered Institute of Personnel and Development, 2001
- [6] Cross, R. and Baird, L., "Feeding organizational memory: improving on knowledge management's promise to business performance", in Cross, R. and Israelit, S. (Eds), *Strategic Learning in a Knowledge Economy: Individual, Collective and Organisational Learning Process*, Boston: Butterworth-Heinemann, 2000.
- [7] Russell, D., Calvey, D. and Banks, M., "creating new learning communities: towards effective e-learning production", *Journal of Workplace Learning*, 15, pp.34-44, 2003.
- [8] Fransman, M., "Information, knowledge, vision and theories of the firm", *Industrial and Corporate Change*, Vol. 3, No. 3, 1994.
- [9] Hiltz, E.S.R., and Wellman, B., "Asynchronous learning networks as a virtual classroom", *Communications of the ACM*, 40(9), 44-49, 1997.
- [10] Imel, S., "Learning Communities/Communities of Practice", *Trends and Issues*, Alert No. 26, 2001.
- [11] Ingrams, P. and Roberts, P., "Friendships among competitors in the Sydney hotel industry", *Australian Journal of Sociology*, Vol. 106 No. 2, pp. 387-423, 2000.
- [12] Khirallah, D.R., "A new way to learn", *InformationWeek*, May, 2000.
- [13] Lave, J. and Wenger, E., *Situated Learning: Legitimate Peripheral Participation*, Cambridge University Press, Cambridge, MA, 1990.
- [14] McCloskey, D.W., Antonucci, Y.L., and Schug, J., "Web-based vs. traditional course development: Identifying differences in user characteristics and performance outcomes", *Proceedings of the International Business Schools Computing Association Annual Conference*, Denver, Colorado, 1998.
- [15] Raffo, C., O'Connor, J., Lovatt, A. and Banks, M., "Attitudes to formal business training and learning amongst entrepreneurs in the cultural industries: situated teaming through 'doing with others'", *Journal of Education and Work*, Vol. 13, pp. 215-30, 2000.
- [16] Swanson, G. and Wise, P., *Digital Futures: Women's Employment in the Multimedia Industries*, Brisbane: Australian Key Centre for Cultural and Media Policy, 1997.
- [17] Tyre, M. and Von Hippel, E., "The situated nature of adaptive teaming in organizations", *Organization Science*, Vol. 8 No. 1, pp. 71-83, 1997.
- [18] Wenger, E., *Communities of Practice: Learning, Meaning and Identity*, Cambridge University Press, Cambridge, MA, 1998.
- [19] Wenger, E.C. and Snyder, W.M., "Communities of practice: the organizational frontier", *Harvard Business Review*, 1 January, 2000.
- [20] Wulf, K., "Training via the Internet: Where are we?", *Training and Development*, 5, 50-55, 1996.

What E-Learning Providers and End Users Should Do Respectively before Initiating Such a Learning Project ¹

Zheng Zhao, Runtian Jing

School of Management, UESTC, Chengdu, 610054, China
fanhuijiaspec@163.com

ABSTRACT

As an E-learning provider, before offering such a product it needs to carry out a thorough analysis in order to understand the customers and trends, and evaluate external environmental forces which include market demand, political and legal forces, social and ethical influences, technology and competition. Through conducting a thorough analysis, a clear strategy can be developed that identifies which customers the e-learning provider is to target and where their product or services will be best positioned. To some extent providers are forced continually to forecast, monitor and assess this environment and to adapt e-learning products or services accordingly before offering learning products. To implement e-learning project successfully, the end users need to understand what e-learning is good at and not so good at. When this basic analysis is complete, one can go into the organization looking for specific learning needs that could be met with e-learning. The advantage of starting by doing is that users will find out where they need help, and where they can manage perfectly well on their own. For any company, the returns must be considered carefully before undertaking such a program. In this article, it also explores the benefits of e-learning investments with a particular focus upon measuring the value of delivering an enterprise e-learning initiative. In addition to the many measurable returns, the paper offers a number of formulas for calculating the ROI and justifying the cost. The article concludes with guidelines detailing the implementation of e-learning initiative.

Keywords: E-learning, Internet technology, distance learning

1. THE BENEFITS OF E-LEARNING

Rather than waiting for a scheduled course to occur, employers and employee now can turn to e-learning programs - available every day, any hour. The "just-in-time" nature of e-learning means specialized learning is available at the moment it's needed without worrying about classroom or instructor availability. The result is that learning is accomplished more rapidly while ensuring content consistency and standardization. In addition, e-learning can streamline the education process by freeing educators' time to focus on learning contents, supplemental information development, and learners' educational needs. E-learning's more efficient use of time can increase productivity and decrease learning costs (Smith, 2001). The benefits of e-learning are as follows.

Efficient learning time. There is minimal work time lost with just-in-time learning. The user can learn anywhere and at anytime at their convenience: self-paced, self-managed, learning-on-demand, distance learning and open learning. The individual can immediately apply what he or she learns to the job when this takes place in the working environment.

Cost-effective learning. E-learning products can be delivered to any networked computer, to a large number of individuals. Programs can be delivered and re-used

with fewer costs than with traditional learning methods. There are no expenses required for the individual's travel, subsistence, or accommodation as with off-site instructor-led courses.

Efficient management. The self-managed learner is a useful resource to the organization, given the right learning and guidance (Pedler et al., 1997). The empowered individuals can organize their own learning methods within the e-learning environment thus freeing up their managers to do other tasks.

Customization. E-learning courses can be customized to meet the organization's specific needs. This flexible learning environment can also be department, product or individual specific.

Knowledge exchange and sharing. Individuals can share their new knowledge and skills with others in the team and department. This increases the comprehension of other employees by integrating and sharing information with other departments, suppliers and customers.

Job performance is increased. Understanding is developed from achieving greater knowledge retention by wanting to learn (Tarr, 1998). Hence, fewer unproductive mistakes are made in the organization. Self-motivated, self-disciplined, self-reliant and self-assessed employees are more autonomous and can

¹ Supported by NSFC (70372032, 79900022)

contribute better to the whole organization performance (Pedler et al., 1997).

There is a lower turnover of staff. Employees are more confident, competent and valued, due to the learning climate and are more likely to stay in the organization (Reay, 1994). Organizations that offer ongoing education have a higher rate of employee retention and benefit from a better-skilled workforce. Organizations are becoming aware of the importance of continual learning.

Acceptance of organization change. Self-directed individuals accept organization change more readily as they adapt by learning new techniques and processes (Reay, 1994). The author French (1999), identifies the self-directed learning benefits to a company by: "... employees who see themselves in a learning environment add value to the companies, increase productivity, are more accepting of change, and can adapt to movement within the company".

Accessibility to current information and new material. Knowledge and skills within an organization can be continually updated through e-learning. The organization can adapt more quickly to the environment in which it operates, due to better-informed employees. According to Steed (1999), skill transfer and retention with e-learning is at least 30 per cent more effective than instructor-led learning.

Utilization of existing resources. A Web browser provides the interface to the Internet and this can access the e-learning environment on any operating platform, such as Windows, UNIX or Mac. A learning program can be delivered to any networked computer without using a different program for different platform.

Consistency and control of learning material and revisions. After the original implementation of an e-learning program that links to Web-based information, further program changes, additions, enhancements and developments can be made on the server where the program resides and everyone in organization can access the latest updated version.

2. E-LEARNING PROVIDERS

E-learning providers are likely to be more successful in attracting customers, if they are much closer to the external environment. Forecasting, understanding and assessing the environment are the greatest challenges and opportunities for e-learning providers. The environment consists of the following uncontrollable factors: market demand, political and legal forces, social and ethical influences, technology and competition.

These factors need to be measured and analyzed using market research and forecasting techniques. The information collected becomes essential to the decision

making of the strategy for an e-learning provider. The factors are discussed next.

2.1 Market demand

The demand for products and services relies on the economic and behavioral aspects of the consumer. A behavioral aspect such as lifestyle trends creates new demands and indicates opportunities for new product development. This influences the direction of an e-learning provider's strategy. An example of lifestyle trends includes individuals working longer hours with greater demands imposed on them, giving less time to keep up-to-date with information and to learn it (Pearn, Roderick and Mulroony, 1995). Also, the geographic locations of an organization increase global business travel for the individuals, which create problems for them about keeping informed, trained and educated. The Internet provides the means for e-learning, not only to satisfy the economic and social change, but also the accessibility to the individual for learning, by promoting a self-directed and flexible learning approach. Demographic changes affect market demand and therefore the strategy. There is an increase in the number of people that are self-employed and working from home, especially in the service sector (Bickerton et al., 1996). The fastest growing group in higher education is part-time students older than 25 (DfEE, 1999). This group tends to seek education to advance their careers, increase their salaries and self-development. For universities and business-to-consumer (B2C) learning providers, these individuals are ideal customers for having e-learning delivered to their home or offices.

2.2 Political and legal forces

The political and legal processes in society affect the way in which an e-learning provider operates. Changes in legal regulations and requirements give rise to many new opportunities and threats, and influence the way in which products and services are offered. Legal requirements that regulate the marketing and advertising products and services over the Internet include: Pricing Regulations, National Standards, Consumer Rights etc.

There are accredited standards being created for e-learning creation. Both e-learning providers and standards committees are recognizing the need to encourage development of Web-based content similar to that of computer software programming. The Sharable Content Object Reference Model (SCORM) standards under development by the United States Department of Defense's Advance Distributed Learning Network (ADL) addresses the need for modular, reusable e-learning content in universities and companies to teach procedures and skills. Other accredited standards include developments from the Institute of Electrical and Electronics Engineers' Learning Technology

Standards Committee (IEEE LTSC); Instructional Management System (IMS) Global Learning Consortium and ARIADNE, a European Union project, that focuses on the creation of standards for e-learning content, management and technology. The IEEE LTSC has also initiated the move of this work to the full International Standards Organization (ISO) standards by establishing ISO Joint Technical Committee 1 (JTC1) Sub Committee 36 (SC36) on Learning Technology.

2.3 Social and ethical Influences

Active involvement by consumer and pressure groups has forced the need for social responsibility and protection of the company's reputation to be important. In order to maintain a good image and reputation among customers, e-learning providers must avoid: misleading ambiguous advertising, inadequate product and service information and hidden financial charges.

The following points can present threats to the organization wanting to provide e-learning:

The e-learning approach can be too innovative for non-technical customers. E-learning needs appropriate understanding and commitment from the customer's management and all those involved in the organization (Gupta et al., 2000).

Organizations can be reluctant to learn over the Internet entirely. Distance-learning courses, such as for further or higher education, where learning is via books, videos and e-mail correspondences has been more accepted by organizations.

Individuals need to understand before accepting any new methods of learning. Decreasing instructor-led learning can make learners anxious. Therefore all affected customers need to be involved at the beginning of the e-learning implementation with the e-learning provider. Also e-learning maybe needs to be introduced into an organization gradually and combined with instructor-led way. This blended learning is more likely to be tolerated by individuals and the organization.

Organizations have a vast amount of knowledge, information and lessons learnt in a wide variety of procedures and in individuals' heads. This diffused knowledge needs to be managed and integrated with the e-learning systems properly (Gupta et al., 2000).

Ethical concerns, including learner privacy, are important to the customer. Internet tools can track Web site visitors, and many customers can provide extensive personal information. This may leave customers open to information abuse if the e-learning providers make unauthorized use of the information in offering their products or exchange electronic lists with other businesses.

2.4 Technology

Internet technology has a direct impact on e-learning providers and end users. With e-learning, the Internet provides the technology element that makes up the product and its delivery as well as providing the distribution channels and a means to promote the e-learning product or services.

Web-based tools and technologies provide the means for designing, developing and delivering e-learning content. The technology includes Web browsers that provide the interface to the Internet, and hypertext markup language (HTML), the standard language that Web pages are written in, that is supported by Web browsers. Web authoring systems assist Web designers to develop the e-learning environments. These consist of either HTML editors or Web page layout programs like Microsoft FrontPage or Macromedia Dreamweaver, and Web development programs like Authorware and ToolBook. Internet tools used can be synchronous or asynchronous communication tools: synchronous tools, including whiteboards, video and audio conferencing, allow individuals to interact in real-time over the Web. Asynchronous tools, including e-mail and forums, allow individuals to learn at any time (French, 1999). Groupware is the technology that allows a group of people to collectively work on shared information and documents over the Web using the synchronous and asynchronous tools (Ravet and Layte, 1997).

Customers can use search engines to find any information regarding the e-learning providers and their e-learning products and services. The e-learning providers need to embed keywords and page titles into their Web pages using the Web authoring systems, in order for the search engines to retrieve a description of an e-learning provider's Web site (Hanson, 2000). However, there are problems affecting search engines. A search will sometimes return thousands of results. Facing this, most users will abandon the search if their desired information is not found in the first or second screen. E-learning providers must therefore make sure their important Web pages are listed within all major search engines and use their Web page content to place important Web pages high on the return list in keywords searches.

Currently there are limitations related to the network capacity of the Internet, the bandwidth, which governs the rate at which information can be transmitted. The reason why video and audio transmission can be relatively slow over the Internet is that the bandwidth cannot cope with the large file sizes (Steed, 1999). Solutions to this are presently being developed by the telecommunication industries. One solution is asymmetric digital subscriber line (ADSL), which uses the existing copper telephone wires. In the meantime a technology called streaming is alleviating the problem for transmitting video, audio and animation sequences,

where an application or file is broken down into small chunks and delivers the beginning of the application while concurrently sending the other compressed chunks.

Security is an important aspect of e-learning and a major area of concern for consumers. If organization learning involves sensitive corporate information, or if payment is necessary over the Internet, then a greater amount of security should be used. The following are the main areas related to Internet security (Hall, 2000; Steed, 1999):

Access control. There are techniques to control which users have access to the information. Corporate Intranets are often protected through the use of firewalls that are gateways between the Intranet and the Internet. The process of authenticating users by issuing them with username and passwords is another type of access control.

Information confidentiality. Encryption is the technology used to ensure information passing through the Internet remains private. It does this by basically scrambling the information from the sender and unscrambling the information for the receiver only. Web browsers and Web servers are being developed with encryption technology built in.

The following points can present technology threats:

Present bandwidth limitations for e-learning means performance for sound, video and complex graphics can be slow, causing long waiting times for these files to download from the Internet, affecting the flow of the learning process.

The initial development time and costs can be significant for authoring customized Web pages for the learning environment and integrating the delivery environment.

Internet connection can be lost or broken due to problems with either the Internet service provider (ISP) or the organization's IT department.

An information infrastructure, i.e. networked PCs, and resources are required within the customer's organization. Therefore, the design, development, implementation, security and management of the e-learning content may need to be maintained by an e-learning provider.

2.5 Competition

Competition is probably the most dynamic of all the environmental forces (Bickerton et al., 1996). A strategy must take into consideration the competitive situation of the environment in which the e-learning provider operates. Successful e-learning providers must satisfy the needs of the customers better than its competitors.

Understanding competitors' strengths and weaknesses, market share and positioning are essential. Careful monitoring and evaluation of the competition, enables the e-learning provider to make better decisions (Kotler et al., 1996).

Not only does the Internet provide access to e-learning for many customers, but also to the competition. It is easy for a competitor to see what other e-learning providers are trying to achieve as well as view their Web site and e-learning products offered. Currently a strong growth and relatively low barriers to entry are attracting newcomers, including competitors from other related business areas such as the consulting industry.

3. E-LEARNING END USERS

Over the last twenty years, the success rate for implementations of technology-based learning has not been high. Huge numbers of companies have spent countless millions on failed initiatives in pursuit of what we now call e-learning.

How can user be sure that it's investment in e-learning will be worthwhile or enduring? Part of the answer lies in the technology. This part is easy to deal with because of global technology developments: The universal adoption of Windows and the html-based web browser, the emerging standards in e-learning, broadband access, networked PCs and better security measures.

How to deal with the rest part of the answer is the main concern of this section.

First, before implementing e-learning, user needs to understand what e-learning is good at and not so good at.

There are skills that can't be learned through an e-learning course. For example, imagine expecting someone to e-learn how to ride a bicycle. User could spend vast sums creating a simulator. Effective? Hardly.

However, user can deliver knowledge and understand more effectively by computer than by any other way, and all skills are enhanced by knowledge and understanding. Achieving this through e-learning frees user to put more energy into practical exercises, role-playing, etc.

So if an organization wants to provide e-learning for its people, and someone tries to offer an online education system, or an exploratory learning system, or a "total solution" that claims to do away with the need for live learning altogether, the managers of the organization must know what to do. Meanwhile managers should not feel they have reduced power from their self-directed, continually learning workforce, as they themselves are also employees of the organization and must also keep their skills update. The manager's emerging role in

e-learning is to act more as facilitator and support the employees through coaching, guiding, mentoring and encouraging continuous learning.

When basic analysis is finished, what should end user do?

1. Go out into organization looking for specific learning needs that could be met with e-learning. If companies do not have the resources to carry out a thorough Needs Analysis, talking to line managers will soon reveal some specific needs. For example, the sales force may need more thorough product knowledge, clerical staff may need to understand how to use computer more effectively.
2. Choose one (or more) urgent need for which e-learning would be suitable. But do not imagine for one second that this is a total e-learning solution. End user needs to train its staff in products, processes and procedures. Without employing specialists, user can start with a very specific topic, or more than one if having the resources, and lay out the information employees need to know and understand.
3. Make it happen. Get one team to do it or employ consultants to do it. The line manager (sales, production, clerical) who identified the pressing learning needs may well be a "leader" and help make it happen. Often a keen manager will find staff members who can contribute knowledge of the subject matter, familiarity with the target learners and their culture, and enthusiasm for the introduction of e-learning. Be realistic in ambitions. Don't aim for perfection or the ultimate in sophistication on first project. Stay practical. Countless organizations are spending more on compiling reports on e-learning, which do not take a single step towards a solution.
4. Evaluate the results. By all means collect learners' reactions to the e-learning. They will be valuable, but do not mislead into thinking that the response to a question about the effectiveness of a course is an actual measure of its effectiveness. E-learning lends itself to paired modules - a learning course and an assessment test. The test is not meant to fail people. It is there to allow them to demonstrate their competence. If everyone achieves a very high score on test, it is perfect. If not, managers should review (a) the questions, in case they are badly worded or needlessly tricky, and (b) the learning content, as there may be topics that are not adequately explained. Remember, however, that the benefits of e-learning are difficult to quantify in terms of changes of behavior and productivity and return on investment.
5. Learn from the experience. Consider what was successful about this pilot. Learn from any failures and shortcomings certainly, but most of all, build on success. User will learn more from this one experience than from five years of reading reports, setting up meetings, and

preparing strategy documents.

6. Calling on the experts for help. Of course, as user carry out e-learning implementation more widely through organization, it may wish to draw on the expertise of specialists in various disciplines. End users cannot escape the e-learning experts entirely, but when anyone tells that one strategy or another is a must-have, user should ask them to explain, without using jargon, exactly what the benefit will be. And if user doesn't really understand the explanation, don't invest in e-learning.

4. CALCULATING THE VALUE OF E-LEARNING

Many companies view learning budget strictly as a cost center, which is susceptible to cuts when the business needs to reduce expenses. For any company, the returns must be considered carefully before undertaking a learning project. In this section, we explore the benefits of e-learning investments with a particular focus upon measuring the value of implementing an e-learning initiative. In addition to the many measurable returns, the section offers a number of formulas for calculating the return on investment (ROI) and justifying the cost.

E-learning is the future wave, but if the wave is ever to come, corporations must prove that it delivers a tangible return on investment. If early adopters integrate methodologies to demonstrate ROI, the wave will change corporate forever.

It is difficult to assign a monetary value to e-learning's benefits. Calculating airline ticket savings because a whole division took an online course is straightforward; but calculating improvements in a manager's interaction with direct reports after an online communications course is more demanding. These improvements are at least as valuable as the travel savings realized by e-learning, but more difficult to measure and justify to senior level decision makers.

Many evaluation experts have developed models to determine the value of specific learning experiences. Any company can follow these formulae to calculate ROI from a learning initiative.

Several ROI models have been developed, but two stand out as practical to implement and be effective in delivery. These two methods, a model by Kirkpatrick (1994) and an enhancement by his colleague, Phillips (1996), form a logical framework to view ROI both from a human performance and business performance perspective. Critical to both is the concept of "chain of effect", which links benefit levels, derived from learning, together. Each level of measurement impacts the previous and next levels. Without accounting for these links, it is difficult to conclude that learning is responsible for any performance improvements.

In his original model, *Evaluating Training Programs: The Four Levels*, Kirkpatrick (1994) considers the value of learning on four levels: participant reaction, skills acquired, skills implemented and organizational benefits accrued. In his book, *Accountability in Human Resource Management*, Phillips (1996) suggests adding another level to Kirkpatrick's model to calculate the return on investment generated by learning. Thus a corporation cannot ultimately measure ROI in isolation; it must base it on measurements gathered at the four preceding levels.

Evaluation is aided by using a question and answer methodology:

Level I – “What are participants’ reactions to the learning and what do they plan to do with the material?” Learners measure this answer with surveys or questionnaires that measure whether the learning was meaningful or enjoyable. These surveys should also include sections on how the employee plans to implement the lessons learned.

Level II – “What skills, knowledge, or attitudes have been changed or acquired and to what extent?” Achievement tests measure how well the employee learned the information or skill presented.

Level III – “Did participants apply what they learned to their jobs?” Observer ratings and observations measure the degree to which the employee applies what he or she has learned. Observers (usually managers and supervisors) must be thoroughly informed in the evaluation system. Managers need to establish a system for leveling out the inconsistencies between observers’ judgment.

Level IV – “Did this on-the-job application produce measurable results?” These results may include increases in productivity and efficiency, decreases in absenteeism and occupational accidents, decreases in customer complaints, etc. Isolating the effects of learning from other variables that produce an effect in these areas, either through statistics or by using a control group, is vital to getting a clear picture of ROI.

Level V – “Did the monetary value of the produced results exceed the cost?” This is the measurement of ROI, which can be calculated in several ways.

The fifth level gets down to quantify the actual return on a company's monetary investment in learning and requires a mathematical formula to determine an answer. Evaluation experts have developed three common formulae for measuring learning ROI, each reflecting a different concept of what a company's return on learning investment can be:

(1) $TACFP - TACNP = PNS$. Subtracting the total administrative costs of the new e-learning program

(TACNP) from those of the former traditional learning program (TACFP) gives the projected net savings (PNS) for learning administration. Although cost savings are certainly important, ROI encompasses much more than just that.

(2) $TCL / \text{number of students} = CPS$. Dividing the total cost of learning (TCL) by the number of students gives the cost per student (CPS) of the learning. This is useful, but again is not a true measure of return on investment. Both formulae, although frequently cited as measures of return on investment, do not measure what monetary value or profit is derived from learning investment.

(3) $TB (\text{in}\$) / TLC = ROI\%$. Dividing the total benefits (TB) of learning in dollars by the total learning program cost (TLC) gives the percentage of ROI in a program. To calculate this formula and get a percentage of ROI, we must determine the tangible benefits of learning and assign a monetary value to benefits such as:

Increase in productivity (units produced, items sold, forms processed, tasks completed, etc.); Improvement in quality (less scrap, less waste, less rework of product, fewer defects, etc.); Reduction in staff turnover; Reduction in lost-time injuries; Reduction in workers' compensation insurance claims; Increase in customer satisfaction as reflected in an increase in repeat sales.

These benefits are often called measurable benefits because the results can be converted easily into a monetary value. Other benefits such as improved communication, enhanced corporate image, improved conflict resolution, increased sensitivity to human diversity, improved employee morale and increased employee loyalty are intangible benefits and more difficult to convert into a monetary value.

With so many benefits to its credit, a company's learning investment should not end up only on the expense side of the balance sheet, ready for budget slicing. By measuring the results of learning carefully and tying learning to strategic metrics used to measure business success, the return on investment from learning can be demonstrated to the company decision makers.

5. GUIDELINES OF IMPLEMENTING E-LEARNING

Many companies have recognized the value of e-learning and view these efforts as an investment in the company's future. Everyone knows a well-educated workforce is vital to a company's success.

5.1 Decision-making

To successfully implement e-learning, first determine how it fits into the organization's needs. Identify needs, articulate how technological tools can help meet those needs, then search for content that's appropriate for the learning's intended purpose. Consider creating a team of

stakeholders to facilitate decision-making that seeks to design a consistent e-learning strategy. To best understand learning needs throughout organization, the team should include representatives from all divisions of the organization.

Create a strategic implementation plan. At first, the organization may have to focus on one of needs rather than tackling every suggested learning issue. Prioritize learning initiatives, create phases and articulate evaluation guidelines to ensure successful implementation from an educational and business perspective.

Once the value of and appropriate use for e-learning are well defined, focus on technological and content-related decisions. First, determine if there's already an e-learning program in place. If e-learning isn't currently in place, articulate necessary performance requirements to guide the search for an appropriate solution.

From a technological perspective, a key component to launch a successful e-learning strategy includes an understanding of learning-management systems and the environment in which e-learning will be carried out. Many e-learning providers offer learning-management systems for different end users. Some are simply technological platforms or structures within which an organization can house its own content, others come with content components.

Before determining the actual technological components, consider content availability. Question each e-learning provider about the kind of content its platform offers and if the content is included in the implementation cost or if it's extra. Also inquire if other content can be added to the platform. These steps will ensure that the platform meets both technological and content requirements.

5.2 Implementation

To generate interest, market the program by demonstrating how employees will benefit from its availability. Provide information explaining the efficiencies and increased productivity that will ultimately be realized from the learning. These efforts will help generate a target around the initiative that'll see it through to a successful implementation.

To augment learning benefits, a blended-learning approach, which combines e-learning with traditional classroom instruction, can be used. The e-learning program can serve as the theoretical foundation upon which participants can build specific skills, either within a classroom or lab setting (Rosenberg, 2001). Such a model capitalizes on the benefits offered by e-learning and optimizes education with face-to-face activities that can accommodate various learning styles, provide social interaction, and focus on communicating.

To create a blended-learning approach, simply determine what tasks can be performed electronically, which ones need a face-to-face or hands-on approach, and how to mix the two. Educational outcomes may remain the same as those articulated for the current construct or modified for the new model. Blended-learning environments may enhance educators' effectiveness, enabling them to use time previously spent delivering basic educational content to focus on specific learning needs of each student.

5.3 Evaluating procedure

Now consider numerous issues about ROI evaluation. Although generally difficult to measure, several direct costs are easy to quantify, such as the number of students, the number of instructors needed. Also calculate the cost of copying handouts, resources such as textbooks, fees charged by the providers, and any travel or room rental expenses.

To begin the evaluation process, quantify the direct and indirect costs for each student. Then quantify costs associated with the educator presenting the learning. The educator's total savings in time and salary is the combined totals of the time and money saved in preparation and teaching. If the educator doesn't spend time preparing for and teaching the basic part of the courses, then he or she can dedicate that time to other educational activities while the learner still acquires the fundamental knowledge via the e-learning solution. Hence, the educator's productivity is increased, effectively doubling his or her educational output. In addition, be sure to include indirect costs necessary to implement e-learning.

The benefits of e-learning include employees' satisfaction level and such business efficiencies as cost and/or time reduction gained by its implementation (Raths, 2001). After calculating total costs and benefits, we can use those formulae mentioned in the previous section to evaluate the e-learning program.

Maybe it is necessary to evaluate whether or not participants are achieving competencies at the same or higher levels than with instructor-led classroom learning.

6. CONCLUSION

The e-learning provider has little control over the external environment in which it operates, and so it is necessary to continually assess this environment and to adapt its e-learning product or services accordingly. Market demand, political and legal forces, social and ethical influences, technology and competition are five main forces affecting e-learning provider's strategy. From conducting a thorough analysis, providers can identify which customers to target and where their product or services will be best positioned.

Before initiating an e-learning project, end user needs to understand what e-learning is good at and not so good at. Doing this will help user find out where it need help, and where it can manage perfectly well on it's own. The paper also explores the greater benefits of an e-learning initiative and offers a number of formulas for calculating the ROI and justifying the cost. At last, the guidelines of implementing e-learning project were discussed.

REFERENCES

- [1]Bickerton, P., Bickerton, M. and Pardesi, U., *Cybermarketing*, Butterworth-Heinemann, Oxford, 1996.
- [2]Department for Education & Employment (DfEE) (1999), *Education and Training Statistics for UK 1999*, 3rd ed., London: The Stationery Office, 1999.
- [3]French, D., *Internet Based Learning*, Stylus Publishing, Sterling, VA, 1999.
- [4]Gupta, B., Iyer, L. and Aronson, J., "Knowledge management - practises and challenges", *Industrial Management & Data Systems*, Vol. 100 No. 1, pp. 17-21, 2000.
- [5]Hall, B., *Web-based Training Cookbook*, Wiley Computer Publishing, Chichester, 2000.
- [6]Hanson, W., *Principles of Internet Marketing*, South-Western College Publishing, Molton, OH, 2000.
- [7]Kirkpatrick, D., *Evaluating Training Programs: The Four Levels*, Berrett-Koehler, San Francisco, CA, 1994.
- [8]Kotler, P., Armstrong, G., Saunders, J. and Wong, V., *Principles of Marketing*, London: Prentice-Hall, 1996.
- [9]Pearn, M., Roderick, C. and Mulroony, C. , *Learning Organisations in Practice*, London: McGraw-Hill, 1995.
- [10]Pedler, M., Burgoyne, J. and Boydell, T., *The Learning Company*, London: McGraw-Hill, 1997.
- [11]Phillips, I. , *Accountability in Human Resource Management*, Butterworth-Heinemann, Oxford, 1996.
- [12]Raths, D., "Measure of Success", *OnLine Learning*. 5(5): 21-26, 2001.
- [13]Ravet, S. and Layte, M., *Technology Based Training*, London: Kogan Page, 1997.
- [14]Reay, D.G., *Understanding the Training Function*, London: Kogan Page, 1994.
- [15]Rosenberg, M., *E-Learning*, New York: McGraw-Hill, 2001.
- [16]Smith, J.: "Blended Learning", Executive Update Online. Available online: <http://www.gwsae.org/Executiveupdate/2001/March/blended.htm>.
- [17]Steed, C., *Web-based Training*, Aldershot: Gower, 1999.
- [18]Tarr, M., "Distance learning - bringing out the best in training", *Industrial & Commercial Training*, Vol. 30 No. 3, pp. 104-6, 1998.

A High Performance XML Querying Architecture

Fangju Wang, Hui Shen

Department of Computing and Information Science, University of Guelph
Guelph, Ontario, Canada, N1G 2W1

ABSTRACT

Data exchange on the Internet plays an essential role in electronic business (e-business). A recent trend in e-business is to create distributed databases to facilitate data exchange. In most cases, the distributed databases are developed by integrating existing systems, which may be in different database models, and on different hardware and/or software platforms. Heterogeneity may cause many difficulties. A solution to the difficulties is XML (the Extensible Markup Language). XML is becoming the dominant language for exchanging data on the Internet. To develop XML systems for practical applications, developers have to address the performance issues. In this paper, we describe a new XML querying architecture that can be used to build high performance systems. Experiments indicate that the architecture performs better than Oracle XML DB, which is one of the most commonly used commercial DBMSs for XML.

Keywords: XML, database, index, query processing

1. INTRODUCTION

In recent years, the needs for Internet data exchange between organizations have grown rapidly, especially in electronic business (e-business). For example, in e-business of B2B (business to business), companies may offer customers and partners to access their databases for information about their products and services. Brokers or consultants may need to process data from databases of many companies for best satisfying client requirements. There has been a trend in e-business to integrate existing stand-alone databases into distributed databases for facilitating data exchange.

XML (Extensible Markup Language) is becoming a major markup language for data exchange on the Internet. The language is especially useful in developing heterogeneous distributed databases, in which databases in different models, and/or on different kinds of platforms are combined into integrated systems. With XML, data from heterogeneous sources can be encoded as XML documents, transmitted on the Internet, and processed together. It has been widely accepted that XML is an indispensable tool in e-business, and that e-business cannot develop well without it.

Performance is a key issue in building a successful XML-based database. The querying techniques for the traditional relational databases were developed to manipulate record sets. High performance of the techniques is achieved by efficient read/write, search, and partitioning of records. However, XML documents are structured as hierarchical trees of nodes, which are essentially different from the table and record structures of the relational data. The querying techniques developed for relational data cannot be directly adopted for XML data. The special features of XML data require new techniques for querying XML data. Currently, research on querying techniques for XML data, including indexing and querying processing,

is still at the preliminary stage. At the time of this writing, there have been no broadly accepted querying techniques XML databases.

In this paper, we describe a new querying architecture, which consists of three indexes, a block parser, and two query processing algorithms. Experiments indicate that the architecture may perform better than Oracle XML DB, one of the most commonly used commercial DBMSs for XML data.

2. RELATED WORK

The related work includes XML data indexing and query processing for XML databases. Most papers about XML indexes focus on building path indexes on XML documents. In [3], an adaptive path index, APEX, is proposed to keep all paths to improve query performance. Ashraf and co-workers proposed two techniques: path tree and Markov tables [1]. They build a path tree to ensure that it fits in memory by deleting low-frequency nodes. Path information is contained in the Markov table.

In the work by Li [6], the index structure is composed of three major components: element index, attribute index and structure index, all of those indexes are based on a numbering scheme. For most index structures for XML data proposed so far, update is a problem because XML element's coordinates are expressed by absolute values. In [5], Kha proposed an indexing structure based on the relative region coordinate that can effectively deal with the update problem.

The existing indexing techniques have major disadvantages. The most severe ones are that the sizes of indexes are very big and the time for building indexes for large XML documents may be very long. For example, in Lore, the total size of the indexes on an XML document of 465 megabytes is 2.2 times larger than the original file size and the time for

building the indexes is about 20460 seconds (5.68 hours) [8].

Much of the work for XML query processing is based on graph theories. Shasha, and co-workers survey both algorithms and applications of tree and graph searching [10]. Prakash Ramanan develops $O(n^2)$ and $O(n^4)$ algorithms based on the concept of graph simulation to minimize tree pattern queries (TPQs) [9]. In the work by Fernandez and Suciu [4], two optimization techniques for queries with regular path expression, both relying on graph schemas, are proposed to restrict search on fragments of graphs and to find the most suitable optimization on all regular queries.

Structural join operations are central to evaluate queries against XML data. In [2], Bruno proposed a holistic twig join algorithm, "TwigStack", for matching an XML query twig pattern. McHugh and Widom created a set of techniques to facilitate XML query processing in the Lore system designed specially for semi-structured data [7].

Currently, the research on querying XML documents is still at the preliminary stage. Some query techniques cannot support very large XML documents [9]. Some have too complex algorithms [9][7]. Some have to work with special indexing systems but the indexing systems are not optimal [4]. Some techniques have to rely on the underlying relational database structures but the join techniques developed for relational data cannot be directly adopted to join XML data.

3. THE INDEXING SYSTEM

As mentioned before, the index system in our approach consists of three indexes. They are path index, element index and block index.

3.1 Path index

The path index contains a set of paths extracted by parsing the schema of the XML document. Every path from the root to a node in the schema tree are identified by a unique number called *PID* (Path ID). Since every element in a document is associated with a node in the schema tree, the PID assigned to the schema node can be assigned to the class of elements in the document associated to the schema node. The path index can be used to find the PID of a given path.

3.2 Element index

The element index is built for finding all the elements associated with a given path. A pair of numbers is assigned to each element's start tag and end tag, which are called StartEID and EndEID respectively. The StartEIDs and EndEIDs are sequential numbers and generated by preorder traversal on the document tree. An ID value is the same as the count of element tags in the whole document. StartEID is the number assigned

to the start tag of an element and EndEID is the number assigned to the end tag. For an element containing string value only, EndEID is the same value as StartEID.

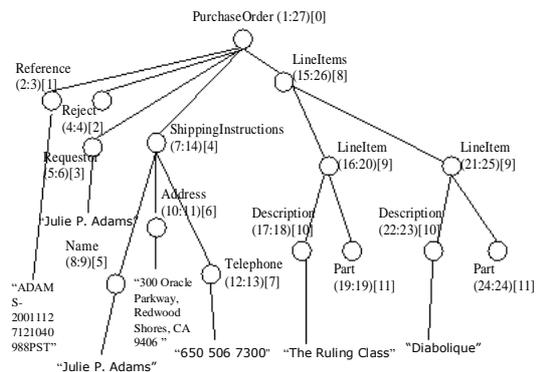


Figure 1. The element numbering strategy

Figure 1 shows an XML document tree and the StartEID and EndEID pairs assigned to the elements. Since every element in a document is associated with a node in its schema tree, the PID assigned to a schema node can be assigned to the corresponding elements in the document. We can search the path in the path index to get its PID, and assign the PID to the element. PID is included in "[]" on every element in the Figure. The StartEID:EndEID pairs with the same PID are stored into a file. The collection of such files is called the element index.

3.3 Block index

The block index is used for identifying the exact file block on disk in which a given element is stored, so that the block can be directly accessed without sequential search. The index contains the StartEID and EndEID pair and PID of the first and last element in every block. A block index is built by recording the information about the tags of the three elements in each block:

1. The first "element with its start tag" in the block.
2. The first "element with its end tag" in the block.
3. The last "element with its start tag or its end tag" in the block.

A tag in XML is always associated with an element. Since an element in the document has an associated PID in the path index, the tags of that element can also be associated with that element's PID. In this paper, a tag's PID refers to its associated element's PID. A tag's EID refers to its associated element's StartEID or EndEID.

The block index plays an important role in achieving the high performance. By comparing a given element's StartEID and ENDEID with the blocks' tag IDs, the block containing the element can be decided and directly read from disk. Then the information about the three elements listed above can be used for block parsing, which will be described in the next section.

4. QUERY PROCESSING ALGORITHMS

In this section, we describe the algorithms based on the indexes for achieving high querying performance.

4.1 Block identification algorithm

The block identification algorithm is used to find the disk blocks containing a given element. The algorithm compares the element's StartEID and EndEID pair with each block's tag information in the block index. In this algorithm, variable *StartBlock* is used to represent the block containing the tag with StartEID and variable *EndBlock* is used to represent the block containing the tag with EndEID. Either *StartBlock* or *EndBlock* is the sequential block number assigned to a block in the block index. The input of this algorithm is the StartEID and EndEID pair of the element to be found, the output is the *StartBlock* and *EndBlock* pair of the block(s) to read. The content of the element crosses over the blocks between *StartBlock* and *EndBlock*.

4.2 Block parser

After the block or blocks containing a given element are read from the disk, we need a special parser to parse the content in the blocks to get the element. This parser is the *block parser*. Along with the block index, the block parser allows an element to be retrieved quickly.

There are two popular XML parsers: SAX and DOM. Both have to scan the entire document from the beginning to end. Scanning XML documents is usually costly. The block parser only parses the contents of the blocks of interests instead of scanning the whole document. It thus helps minimize the I/O costs in retrieving elements.

In the block parser, an *EID counter* is defined to trace the element ID assigned to every element's tag and a *step stack* is defined to trace the element's name in the block. The block parser initializes the EID counter and the step stack by the tag's information recorded in block index.

The blocks to be parsed are a sequence of blocks between *StartBlock* and *EndBlock*. The parsing is started from the first character in the first of those blocks. When the block parser reads a start tag, it pushes the element's name in that tag as a step onto the top of the step stack. When the block parser reads an end tag, it pops up the step from the top of the step stack. There is an exception: when the first tag is a start tag, block parser ignores it and does nothing.

A path, here called *stack path*, can be built by connecting the steps from the bottom to the top in the step stack. Every time after pushing an element's name to the step stack, the block parser builds a stack path and looks up for PID in path index for the stack path. Here, the PID is called *stack PID*. If the stack PID is

among the PIDs of the source nodes in the query trees, the block parser stores the content between the current start tag and its matching end tag into a buffer in main memory or to temporary files on disk.

4.3 Single-root query algorithm

We categorize XML queries into the *single-root queries* and *multi-root queries* and apply different algorithms to process them. In the following discussion, the queries are represented in the Xquery language. A query has a For clause, a WHERE clause, and a RETURN clause.

In a single-root query, there is only one variable defined in the FOR clause that can serve as the root of all the paths referenced in the query. *Source elements*, which are elements referenced in the WHERE clause, must be processed in the scope of root elements. Source elements in different root element scopes cannot be processed together. The result of the operation can be generated within the scope of every root element.

First of all, the PID assigned to the root node is used to obtain the first StartEID and EndEID pair from the element index. The first root element's scope can be built by the StartEID and EndEID pair. Then, the PID assigned to every source element is used to read the corresponding element index. By an algorithm for determining the "ancestor-descendent" relationships, every source element's StartEID and EndEID pair is compared with the root element's scope and then all source elements' StartEID and EndEID pairs in the current root query element's scope are decided.

Through the "Block Identification Algorithm", the exact blocks containing a desired source element in the WHERE clause can be identified. The content of these blocks can be read and parsed by the block parser to fetch the content of that element to a buffer. When all the source elements in the WHERE clause are fetched, the query condition is evaluated. When a set of source elements satisfies the query condition, the blocks containing the associated elements in the RETURN clause are identified. The content of the blocks are parsed and values of those elements are output as the result immediately. Figure 2 is the diagram illustrating this algorithm.

and 600 megabytes, on a schema of “purchasing orders”. For each document, we examine the total size of indexes and the time for creating indexes in both our system and Oracle XML DB. A set of queries are applied to the data. The queries are used to perform operations of selection, equal join, θ -join on the data. The time required by both systems are recorded and compared.

5.2 Experimental results

5.2.1 Building indexes

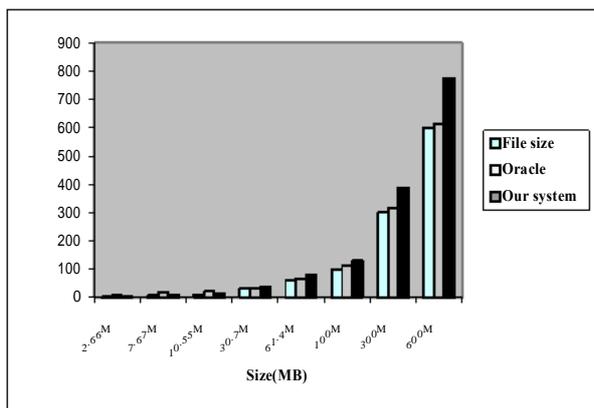


Figure 4 Space used by the two systems.

A path index is a memory structure that is generated by parsing the schema. Generally the size of a schema is very small, so we ignore the time for parsing a schema to create the path index. The time spent on building the other two indexes is significantly affected by the time for parsing the document and the time for writing element index and block index to disk.

Figures 4 and 5 gives the comparison of the space usage in Oracle and our system, and the time required by the two system to organize data and create indexes. Our system needs 20% to 40% of the time by Oracle XML DB and more space when data sizes are big.

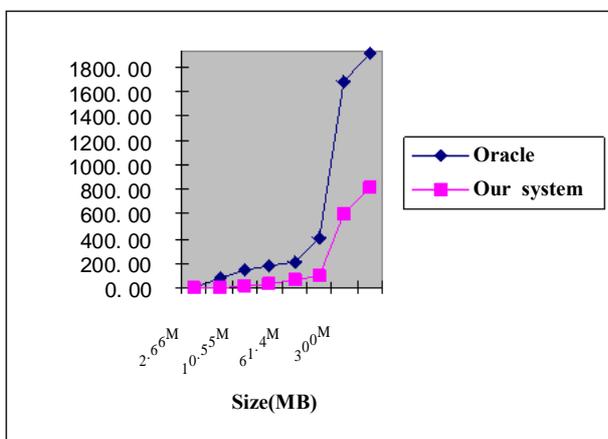


Figure 5. Time required by the two systems.

5.2.2 A query of selection

The following query is designed to select an element and retrieve its parent as well as all the children of the parent element. This is a single-root query.

```
For $c in document("PurchaseOrder.XML")/PurchaseOrder/
Where $c/Reference="ADAMS-20011127121040988PST"
```

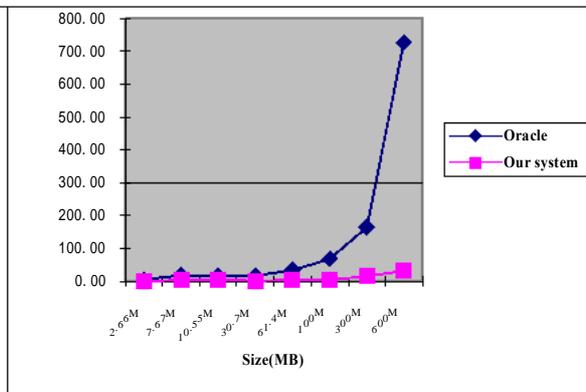


Figure 5. Comparison of the time for executing the selection query.

Figure 5 illustrates the response time for evaluating the selection query with XML documents of different sizes in both Oracle XML DB and our system. The results indicates our system is faster than Oracle XML DB when executing such a query.

5.2.3 A query of document join

The following is a query for joining two XML documents. The WHERE clause includes an equality condition. Element pairs satisfying the condition are selected and the number of the elements in one of the classes are returned as query results.

```
For $c in document("PurchaseOrder1.XML") /Purchase
Order/,
$d in document("PurchaseOrder2.XML")/Purchase
Order/ShippingInstructions
Where $c/Requestor=$d/Name
Return DISTINCT({$c/Requestor})
```

Figure 6 shows the experimental results in testing this query. The time required by our system is basically proportional to the document size. For most of the documents, our system performs better than Oracle XML DB. Since we have no knowledge about how exactly Oracle handles the data, we are unable to explain the behavior of Oracle.

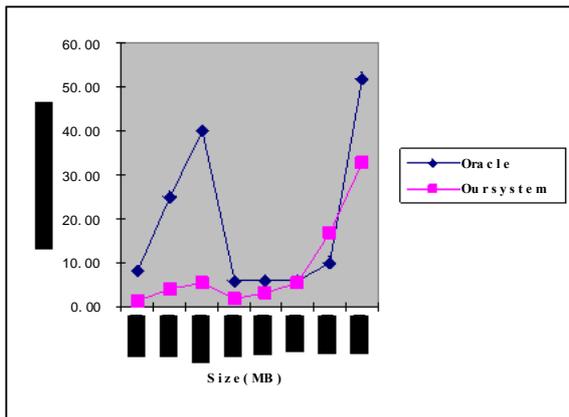


Figure 6. Comparison of the required for executing the join query.

6. CONCLUSION

In this research, we develop a novel architecture for querying XML data. The core of this architecture includes the block index and block parser, which overcome disadvantages of the existing general purpose indexing systems: large sizes and long creation time. Based on the indexing system, a set of algorithms can achieve high performance in evaluating queries. Experiments indicate that the system implementing the architecture requires less time to create the indexes than the time by Oracle XML DB to load the same files. Our system is 2 to 12 times faster than Oracle XML DB. For most of the test data, our system is faster than Oracle XML DB in executing the same queries. Our system requires only 0.5 % to 50% of the time required by Oracle XML DB to evaluate the same test queries.

REFERENCES

1. A. Aboulnaga, A. R. Alameldeen and J.F. Naughton, 2001, "Estimating the selectivity of XML Path Expressions for Internet Scale Applications", *Proceedings of the 27th VLDB Conference*, Roma, Italy.
2. N. Bruno, N. Koudas and D. Srivastava, 2002, "Holistic Twig Joins: Optimal XML Pattern Matching", *ACM SIGMOD 2002*, Madison, Wisconsin, USA.
3. C. W. Chung, J. K. Min and K. Shim, 2002, "Apex: An Adaptive path index for XML data", *ACM SIGMOD '2002 June 4-6*, Madison, Wisconsin, USA.
4. M. Fernandez and D. Suciu, 2002, "Optimizing Regular Path Expressions Using Graph schemas", *Proceedings of IEEE 14th International Conference on Data Engineering*.
5. D. D. Kha, M. Yoshikawa and S. Uemura, 2001, "An XML Indexing Structure with Relative Region Coordinate", *17th International Conference on Data Engineering*.
6. Q. Li and B. Moon, 2001, "Indexing and Querying XML Data for Regular Path Expressions", *Proceedings of the 27th VLDB Conference*, Roma, Italy.
7. J. McHugh and J. Widom, 1999, "Query Optimization for XML", *Proceedings of 25th VLDB Conference*, pp315-326, Edinburgh, Scotland.
8. L. K. Poola and J. R. Haritsa, 2001, "SphinX: Schema-conscious XML Indexing", Database Systems Laboratory, Dept. of Computer science & Automation, Indian Institute of Science, Bangalore.
9. P. Ramanan, 2002, "Efficient Algorithms for Minimizing Tree Pattern Queries", *In ACM SIGMOD 2002 June 4-6*, Madison, Wisconsin, USA.
10. D. Shasha, J. T.L. Wang and R. Giugno, 2002, "Algorithmics and Applications of Tree and Graph Searching", *Proceedings of the 21st ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, Madison, Wisconsin.

Application of Weighted Support Vector Machines to Network Intrusion Detection

Yinshan Jia^{1,2}, Chuanying Jia², Hongwei Qi³

¹School of Information Technology, Liaoning University of Petroleum and Chemical Technology, Fushun 113001, China

²Dalian Maritime University, Dalian 116026, China

³Fushun Ethylene Chemical Co. Ltd., Fushun 113004, China
{jiayinshan, chuanyingjia, qi_hongwei}@sina.com

ABSTRACT

Support Vector Machines(SVMs) have succeeded in many classification fields. Some researchers have tried to apply SVMs to Intrusion Detection recently and got desirable results. By analyzing C-SVM theoretically and experimentally, we found that C-SVM had some properties which showed C-SVM was not most suitable for Network Intrusion Detection. First, C-SVM has different classification error rates on different classes if the sizes of training classes are uneven. Second, C-SVM is over-dependent on every training sample, even if the samples are duplicated. Third, C-SVM does not make a difference between training samples. According to these characteristics of C-SVM and the fact that the size of network normal data is always much larger than that of intrusion data and the fact that the importance of attack data is different from each other, an extended C-SVM, termed weighted C-SVM is proposed in this paper. Weighed C-SVM introduces two parameters, class weights and sample weights. Class weights are used to adjust false negative rate and false positive rate of each intrusion class. And sample weights are used to emphasize importance of some intrusion samples. Experiments showed that Weighted C-SVM was more effective than C-SVM in network intrusion detection systems.

Keywords: Network Intrusion Detection, Support Vector Machines, Weighted Support Vector Machines, Machine Learning

1. INTRODUCTION

With the rapidly increasing connectivity and accessibility to the Internet, network security has been paid more and more attention to. Intrusion detection systems are considered an effective measure against network attacks. Intrusion is generally defined as the violating confidentiality, integrity and availability of computer or computer network system[1]. The kernel of intrusion detection is to extract the behavior model of networks. Currently, many methods have been applied to create detection model, including Neural Networks, Data Mining, and so on[2]. All these methods learn on training data sets first to create detection model, and then use the detection model created by learning process to monitor future behaviors of network. Intrusion detection is a classification problem in nature. It classifies the network behavior to normal class and attack class, or normal class and attack classes. The previous is a binary classification problem and the other is a multi-class classification problem.

Support vector machines are canonized by many researchers and have been applied to many classification fields successfully. The advantages of SVMs over conventional classification methods are its high generalization ability especially if the number of training data is small, its adaptability to various classification problems by changing kernel functions, and its global optimal solution. Recently, some researchers began to apply SVM technology to intrusion

detection. Mukkamala, et al.[3] described approaches to intrusion detection using support vector machines and demonstrated that efficient and accurate classifiers can be built to detect intrusions using SVMs. They also compared the performance of neural networks based and SVM based systems for intrusion detection, and concluded that at the same level of accuracy, the training time for SVMs was significantly shorter, and running time of SVMs was also notably shorter. They also used SVMs in feature ranking and selection for intrusion detection systems[4] and identifying key variables for intrusion detection [5]. In these papers, they did only binary classification experiments, and got very desirable results. John S. Baras and Maben Rabi[6] used SVMs to detect intrusion on a host computer system. In order to handle variable-length strings, the approach combined the ability of an HMM generative model. Ming Luo, et al.[7] and Fugate, et al. [8] used unsupervised clustering and support vector machines to anomaly detection. Other works on applications of support vector machines on intrusion detection can be found in[9],[10],[11],[12].

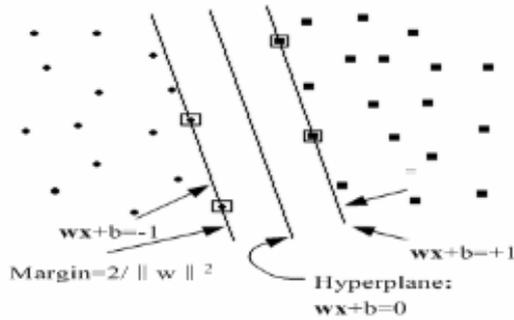


Figure 1. Principle of support vector machines

2. SUPPORT VECTOR MACHINES

Support Vector Machines, proposed by Cortes and Vapnik[13][14], is a relative new machine learning methodology based on statistical learning theory. This support vector machine is considered as the standard support vector machine and C-SVM[15][16]. Its basic idea is to find the hyperplane which can separate data belonging to two classes with maximum margin. This hyperplane is called optimal hyperplane. Figure 1 shows the principle of SVMs.

In Figure 1, the points marked with rectangles are Support Vectors (SVs), which determine the hyperplane and the margin. In the case of nonlinear separation, SVMs use a map $f : X \rightarrow H$ to transform each data point x to $f(x)$ in higher dimensional feature H so that the data points can be separate linearly more probably. X is called input space, and H is called feature space. If $f(x)$ can not separate linearly in H , then SVMs find a hyperplane in H that minimize the error cost. This kind of hyperplane is called optimal hyperplane.

2.1 C-Support Vector Machine

Given $Z = \{(x_i, y_i) | x_i \in \mathbb{R}^n, y_i \in \{+1, -1\}, i = 1, 2, \dots, m\}$ set of samples, where x_i is a data vector, y_i is the label of the class that x_i belongs to. In order to seek the hyperplane that best separates the two classes from each other with the widest margin, we need to solve the following optimization problem[14]:

$$\min_w \psi(x) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i, \tag{1}$$

$$\text{s.t. } y_i \langle w, x_i \rangle + b \geq 1 - \xi_i \tag{2}$$

$$\text{and } \xi_i \geq 0, i = 1, \dots, m. \tag{3}$$

Where w is a normal vector of the hyperplane $\langle w, x \rangle + b = 0$, b is the bias of the hyperplane, C is a penalty factor that determines the tradeoff between the maximization of the margin and the minimization of error cost, $\langle w, x \rangle$ denotes dot product of w and x , each ξ_i is a slack variable that denotes the distance from x_i to margin plane $\langle w, x \rangle + b = y_i$. In order to separate data more precisely, we use a map $f : X \rightarrow H$,

where H is a higher dimensional space. Taking into account the compute complexity, we select a kernel function $k(x, x') = \langle \phi(x), \phi(x') \rangle$. Introducing Lagrange multipliers a_i and b_i , we have:

$$L(w, x, b, a, b) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m a_i x_i - \tag{4}$$

$$\sum_{i=1}^m [a_i (y_i (\langle w, f(x_i) \rangle + b) - 1 + x_i) + b_i x_i].$$

This function has to be minimized with respect to the primal variables w, x, b , and maximized with respect to the dual variables a_i, b_i . To eliminate the former, we compute the corresponding partial derivatives and set them to 0, obtaining:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^m a_i y_i f(x_i) = 0, \tag{5}$$

$$\frac{\partial L}{\partial b} = - \sum_{i=1}^m a_i y_i = 0, \tag{6}$$

$$\frac{\partial L}{\partial x_i} = C - a_i - b_i = 0. \tag{7}$$

Substituting (5), (6) and (7) into L , using $a_i, b_i \geq 0$, and incorporating kernels for dot products, we obtain the dual Lagrangian:

$$\max_w (a) = \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j=1}^m a_i a_j y_i y_j k(x_i, x_j), \tag{8}$$

$$\text{s.t. } 0 \leq a_i \leq C, i = 1, \dots, m, \tag{9}$$

$$\text{and } \sum_{i=1}^m a_i y_i = 0. \tag{10}$$

where $a_i, i=1, \dots, m$, are Lagrange multipliers. If $a_i > 0$, then x_i lies on or in margin, and x_i is called a standard support vector. If $0 < a_i < C$, then x_i lies on margin, and x_i is called a in-bound support vector. If $a_i = C$, then x_i lies in margin or misclassified, and x_i is called a bounded support vector[15]. By solving the above dual Lagrangian, we obtain:

$$w = \sum_{i=1}^m a_i y_i f(x_i). \tag{11}$$

To Compute b , we take into account that due to KKT conditions, for in-bound support vectors for which $\xi_i = 0$, constrain (2) becomes

$$\sum_{i=1}^m y_i a_i k(x_i, x_j) + b = y_j. \tag{12}$$

Thus, b can for instance be obtained by averaging

$$b = y_j - \sum_{i=1}^m y_i a_i k(x_i, x_j). \tag{13}$$

over all data with $a_j > 0$ [15]. The resulting decision function can be shown as

$$f(x) = \text{sgn}\left(\sum_{i=1}^m a_i k(x_i, x) + b\right). \quad (14)$$

From (14), we know that using SVMs to classify new data need not to compute w after training SVMs. In most cases, we don't know the map f . After training SVMs, we store kernel function k_{x_i} which is nonzero and corresponding y_i, x_i , or store kernel function k_{a_i, y_i} and corresponding x_i . When a new data point needs to be classified, we can use these stored values and decision function (14) to classify it.

2.2 Analysis of C-SVM

In C-SVM, the KKT condition is:

$$a_i(y_i(\langle w, x_i \rangle + b) - 1) - \xi_i = 0 \quad (15)$$

$$b_i x_i = (C - a_i) x_i = 0. \quad (16)$$

Therefore, there are three cases as follows:

- 1) If $a_i = 0$, according to (16), then $x_i = 0$. x_i is correctly classified.
- 2) If $0 < a_i < C$, according to (15) and (16), then $y_i(\langle w, x_i \rangle + b) - 1 = \xi_i > 0$ and $x_i = 0$. x_i is a in-bound support vector.
- 3) If $a_i = C$, according to (15) and (16), then $y_i(\langle w, x_i \rangle + b) - 1 = \xi_i = 0$ and $x_i \neq 0$. x_i is a bounded support vector. If $\xi_i = 1$, x_i is correctly classified. If $\xi_i > 1$, x_i is misclassified.

Suppose N_{BSV+}, N_{BSV-} are the number of bounded support vectors in positive class and negative class respectively, and N_{SV+}, N_{SV-} are the number of support vectors in positive class and negative class respectively, and m_+, m_- are the number of data points in positive class and negative class respectively. According to (10), we have:

$$\sum_{i=1}^m a_i = \sum_{i=1, y_i=+1}^m a_i + \sum_{i=1, y_i=-1}^m a_i. \quad (17)$$

$$\sum_{y_i=+1} a_i = \sum_{y_i=-1} a_i. \quad (18)$$

$$N_{BSV+} \leq \sum_{y_i=+1} a_i. \quad (19)$$

$$N_{BSV-} \leq \sum_{y_i=-1} a_i. \quad (20)$$

Incorporating (19) and (20), we have:

$$N_{BSV+} \leq \sum_{y_i=+1} a_i \leq C N_{SV+}. \quad (21)$$

Similarly, we can obtain:

$$N_{BSV-} \leq \sum_{y_i=-1} a_i \leq C N_{SV-}. \quad (22)$$

Dividing (21), (22) by $C N_{SV+}$ and $C N_{SV-}$ respectively, and supposing $\sum_{y_i=+1} a_i = A$, we obtain:

$$\frac{N_{BSV+}}{m_+} \leq \frac{A}{C m_+} \leq \frac{N_{SV+}}{m_+}. \quad (23)$$

$$\frac{N_{BSV-}}{m_-} \leq \frac{A}{C m_-} \leq \frac{N_{SV-}}{m_-}. \quad (24)$$

From (23) and (24), we know that:

- 1) If $m_+ = m_-$, the fraction of bounded support vectors and fraction of support vectors in positive class is equal to that in negative class.
- 2) If $m_+ > m_-$, the fraction of bounded support vectors and fraction of support vectors in positive class is less than that in negative class.
- 3) If $m_+ < m_-$, the fraction of bounded support vectors and fraction of support vectors in positive class is greater than that in negative class.
- 4) Omitting any data point, even if it is a duplicated data point, will influence the result of classification.

Besides the four above characteristics, we can know intuitively C-SVM has other characteristics as follows:

- 5) If the training set has duplicated samples, there would be more duplicated support vectors. In this case, classification will be slower.
- 6) C-SVM does not take different importance of training samples into account. This would result in misclassification of important samples.

According to the characteristics analyzed above, we can conclude that C-SVM is not most suitable to intrusion detection. In intrusion detection, the training sets generally contain more normal samples than attack samples and contain many duplicated samples. New SVMs overcoming the shortcomings of C-SVM are needed.

3. WEIGHTED C-SUPPORT VECTOR MACHINE

In this section, we propose a novel C-SVM, termed weighted C-SVM, WC-SVM for short. The primal question in WC-SVM is

$$\min_{w, \xi, \rho} (w, \xi, \rho) = \frac{1}{2} \|w\|^2 + \mu \sum_{i=1}^m \xi_i, \quad (28)$$

$$\text{s.t. } y_i(\langle w, x_i \rangle + b) \geq -\xi_i, \quad (29)$$

$$\text{and } \xi_i \geq 0, i = 1, \dots, m. \quad (30)$$

Where $\mu_{y_i} \geq 1$ is the weight factor of class y_i , $\mu_{y_i} > 0$ is the important factor of data point i . The bigger μ_{y_i} is, the more important i is. The primal question in WC-SVM is same as that of C-SVM if each μ_{y_i} is specified to 1. Comparing (1) and (28), we can understand the meaning of $\sum \xi_i$ in (28). $\sum \mu_{y_i} \xi_i$ in (28) is practical cost of errors, and that $\sum \xi_i$ in (1) is geometric cost of errors.

With method same as that in C-SVM, we can obtain dual Lagrangian:

$$\max W(a) = \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j=1}^m a_i a_j y_i y_j k(x_i, x_j) \quad (31)$$

$$\text{s.t. } 0 \leq a_i \leq C \mu_{y_i}, i = 1, \dots, m. \quad (32)$$

$$\text{and } \sum_{i=1}^m a_i y_i = 0. \tag{33}$$

The decision function is

$$f(x) = \text{sgn}(\sum_{i=1}^m y_i a_i k(x, x_i) + b). \tag{34}$$

With analysis method same as in C-SVM, we have:

$$\frac{N_{BSV+}}{m_+} \# \frac{A}{C \sum_{y_i} \xi_i \eta_i} \frac{N_{SV+}}{m_+}. \tag{35}$$

$$\frac{N_{BSV-}}{m_-} \# \frac{A}{C \sum_{y_i} \xi_i \eta_i} \frac{N_{SV-}}{m_-}. \tag{36}$$

We can control the upper bound on the fraction of margin errors and fraction of SVs of each class by tuning the parameters η_i and ξ_i in (35) and (36). We think this is useful and important in network intrusion detections. First, attack data is always much less than normal data in sample set. If C-SVM is used, error rate on attack data will be bigger than that on normal data. This is not expected if the attack data contains some data whose corresponding behavior is very harmful to network systems. Second, detection is needed to be real time to avoid attack. We know from (14) and (34) that the number of SVs influences the detection speed. Although C-SVM provides a mean to tune the tradeoff between fraction of margin errors and fraction of SVs, the mean tunes the tradeoff of attack data and normal data simultaneously. In practice, we prefer tuning the tradeoff of each class independently to tuning the tradeoff of each class simultaneously. WC-SVM has advantages over C-SVM in these aspects apparently.

Moreover, attack behaviors do different harms to network systems. Some behaviors harm the systems seriously, and others do slightly. Some new behaviors are very harmful, and we expect IDS can detect them as long as the training sample set contains the data corresponding to them. If the data corresponding to the new harmful behaviors is very sparse in training set, C-SVM would misclassify them. This will result in false negatives in IDS. We introduce an important factors every train sample in WC-SVM in order to handle this case. If some data in training set is important, we can assign a big value to their corresponding η_i otherwise, we can assign a small value to their corresponding η_i .

4. EXPERIMENTS

In this section, we use some experiment results to compare C-SVM and WC-SVM. Experiments on C-SVM were done with libsvm[17], and experiments on WC-SVM were done with a new program based on libsvm. Training samples and testing samples were selected randomly from KDD'99 data[18]. The kernel function used is RBF. Results of experiments are shown in Table 1, Table 2 and Table 3.

In Table 1, MR,NC,AC,NS denote Margin Errors, Normal Class, Attack Class and the Number of SVs

respectively. In Table 2, FNR,FPR denote False Negative Rates and False Positive Rates respectively. In Table 3, NCAS, NMNST, IFNS denote the Number of Changed Attack Samples, the Number of Misclassified New Samples In Training, Important Factor of New Samples respectively.

Table 1. Results of training experiments

Experiment No.	1	2	3	
Number of Attack Data	1000	100	500	
Number of Normal Data	1000	2000	3000	
C-SVM	C	1000	1000	1000
	Time(Seconds)	3.30	1.48	183.57
	ME of NC	0	14	0
	ME of AC	103	18	103
	NS in NC	196	115	232
WC-SVM	NS in AC	118	47	119
	C	1000	1000	1000
	μ_{+1}	1	1	1
	μ_{-1}	10	20	50
	s_i	1	1	1
	Time(Seconds)	65.51	3.645	233.57
	ME of NC	120	50	338
	ME of AC	0	0	0
	NS in NC	215	168	463
	NS in AC	43	21	36

Table 2. Results of test experiments

Experiment No. of Model	1	2	3	
Number of Test Samples	2000	2000	2000	
C-SVM	Time(Seconds)	2.53	1.16	2.27
	FNR(%)	100	75.60	100
	FPR(%)	0.19	8.34	0.20
WC-SVM	Time(Seconds)	1.64	1.39	3.45
	FNR(%)	93.40	25.32	37.1
	FPR(%)	13.04	56.60	36.54

Table 3. Results of test experiments on samples with important factors

Experiment No.	1	2	3	
NCAS	10	15	20	
Number of Test Samples	2000	2000	2000	
C-SVM	NMNST	2	2	1
	IFNS	10	20	100
WC-SVM	NMNST	1	0	0

From results of experiments, we know that WC-SVM provide us means to adjust the correct rate and error rate of classification of each class and to emphasize the importance of each training sample.

5. CONCLUSIONS

We have presented a new SVM termed WC-SVM. From the point of application, WC-SVM has advantages over C-SVM. We can not only specify a weight factor for each training set but also specify an importance factor for each data point. If a data point is misclassified, the penalty error in WC-SVM equals the product of the penalty error in C-SVM and the importance factor of this point and the weight factor of the class which this point belongs to. This results in lower misclassification rates of training classes which have bigger weight factors and results in less probability of misclassification of training samples which have bigger important factors. In our opinion, this idea is useful to

intrusion detection. This is because we always need different classification performance, and we need to run intrusion detection systems before we collect much attack data.

REFERENCES

- [1]Bace, R. G., Intrusion Detection, Macmillan Technical Publishing, 2000
- [2]David,J.M.,Computer Intrusion Detection and Network Monitoring:A Statistical Viewpoint, Springer-Verlag, 2001.
- [3]Srinivas Mukkamala, Guandalupe Janoski, Adrew Sung, "Intrusion Detection Using Neural Networks and Support Vector Machines", Proceedings of IEEE International Joint Conference on Neural Networks 2002, pp1702-1707, Hawaii, May 12-17, 2002
- [4]Srinivas Mukkamala and Andrew, H.Sung, "Feature Ranking and Selection for Intrusion Detection Systems Using Support Vector Machines", Proceedings of the 2002 IEEE International Conference on Information and Knowledge Engineering, pp503-509, Las Vegas, Nevada, June 24-27,2002
- [5]Srinivas Mukkamala, Andrew,H.Sung and Ajith Abraham, "Identifying Key Variables for Intrusion Detection Using Soft Computing Paradigms", Proceedings of the 2003 IEEE International Conference on Fuzzy Systems, St. Louis, Missouri, USA, May 25-28,2003
- [6]John, S. Baras and Maben Rabi. "Intrusion Detection with Support Vector Machines and Generative Models", Research Report, University of Maryland, 2002
- [7]Luo Ming,Wang Lina,Zhang Huanguo and Chen Jin, "A Research on Intrusion Detection Based on Unsupervised Clustering and Support Vector Machines", ICICS2003,LNCS 2836, pp325-336, Springer-Verlag , 2003
- [8]Mike Fugate and James, R.G., "Normally Detection Enhanced Classification in computer Intrusion Detection", SVM 2002, LNCS 2388, pp.186-197, Springer-Verlag , 2002
- [9] Kim Dong Seong and Park Jong Sou, "Network-Based Intrusion Detection with Support Vector Machines", ICOIN 2003, LNCS 2662, pp 747-756, Springer-Verlag , 2003
- [10]Rao Xian, Dong Chunxi and YANG Shaoquan, "An Intrusion Detection System Based on Support Vector Machine"[In Chinese], Journal of Software, Vol. 14, No. 4, pp798-803, 2003
- [11]Li Hui, Guan Xiaohong, Zan Xin and Han Chongzhao. "Network Intrusion Detection Based on Support Vector Machine" [In Chinese]. Journal of Computer Research and Development, Vol. 40,No. 6,pp799-807, 2003
- [12]Tran Quang-anh, Zhang Qianli and Li Xing. "SVM Classification-based Intrusion Detection System"[In Chinese], Journal of China Institute of Communications, Vol. 23, No. 5,pp51-55,2002
- [13]Cortes, C. and Vapnik, V., "Support vector networks", Machine learning, Vol. 20,No. 3, pp273-297,1995
- [14]Vapnik, V., The nature of statistical learning theory, Springer-Verlag, 1995
- [15]Schölkopf, B. and Smola, A. J., Learning with kernels, MIT Press, 2002
- [16]Schölkopf, B., Smola, A. J., Williamson, R. C., et al., "New support vector algorithms", Neural Computation, Vol. 12,No.5, pp1207-1245, 2000
- [17]Chang Chih-Chung and Lin Chih-Jen, "LIBSVM: a library for support vector machine, 2001", <http://www.csie.ntu.tw/~cjlin/libsvm>
- [18]<http://kdd.ics.uci.edu/databases/kddcup99/task.htm>

Customers Confidence in E-Business: An Evaluation of Australian Practices A Case Study

Daniel Chandran

Faculty of Information Technology, University of Technology, Sydney, Australia
noble@it.uts.edu.au

ABSTRACT

Many Internet users perceive that there is a large risk to their privacy and security when they buy products and services or submit personal information online. Although the perception of risk may be greater than the actual risk, it is still a cause for concern. An e-business must address customers' perceived risk just as much as any actual risks. In other words, the issue of trust is fundamental as security and privacy issues are of major concern for many users. Customers are looking for security policies and procedures businesses are implementing and their responsibilities in keeping customer information secure.

This paper deals with trust e-businesses can create through their websites. A comprehensive set of criteria had been used to evaluate Web sources with a high level of meticulousness. Criteria include accuracy, authority, currency, scope, and relevance. A methodology has been developed and two main industries actively operating in e-business within Australia, namely, the Banking and Retail industries have been chosen. Under each industry, two businesses were identified and sources evaluated. It is presumed that using such criteria would help to gain customers confidence. Findings are reported.

Keywords: e-commerce, Privacy and Security, Trust, Evaluation Criteria.

1. INTRODUCTION

The Internet provides instant and universal access to a range of business sites. Any business that offers services through the Internet needs to be concerned about network security as it is being constantly attacked. A computer network system is exposed to potential threats from anywhere on the public network. Businesses need to protect against such threats and implement stringent security measures. Many Internet users perceive that there is a large risk to their privacy and security when they buy products and services or submit personal information online. Although the perception of risk may be greater than the actual risk, it is still an issue for concern. An e-business must address customers' perceived risk just as much as any actual risks. From the point of customers, they are concerned about the hackers who can gain access to steal their information.

Trust can be defined as faith or confidence in a person or organisation. The level of security depends on the nature of business and the risks involved. For example, a web site providing information on the products of a company may not require the same level of security as an online banking Website. An e-business must identify the security issues that would provide the business more confidence and better relationship among its customers.

In the last few years, e-commerce went through difficult path of decline. One of the reasons is that the customers were not sure about the design and security holes. E-business in Australia is widespread and delivering benefits to the national economy. This paper is written from the customers' point of view and tries to list the

ways through which the customers can evaluate the sources in terms of security and privacy before they can conduct any business online. Some frequently asked questions are:

- § How is the data protected once it is delivered to the e-business?
- § How are credit card transactions authenticated and authorized?
- § Can I trust this Website?
- § How can I best evaluate e-business information provided on the Internet?

When the customers are satisfied with the security measures provided by the business, they make frequent return visits to the web site or make repeat purchases from the web site. If they are not confident about the policies, the situation could change. Several customers research about a product on the website and purchase it offline. Customers may not understand the security policies described in the websites. In such a situation, how can they come to a conclusion whether a particular site is secure? It is important that users undertake a systematic approach to select sources, which they consider as secure. To exchange information through the use of technology, we must be certain that it is as safe as exchanging information face-to-face. In this study, numerous web searches were carried out to gather and select sources pertaining to businesses that offer online business in Australia and the information they provide to their customers.

A range of measures is required to inculcate confidence in customers to conduct business over the Internet.

Customers may have to evaluate sources with a high level of meticulousness, by using a comprehensive set of criteria. Criteria include Accuracy, Authority, Currency, Scope, and relevance. A methodology has been developed for this research and two main industries actively operating in e-business within Australia, namely, the Banking, and Retail industries identified. Under each category, two businesses were chosen and the sources evaluated. It is presumed that using such a set of criteria would help customers to gain confidence to choose and use a web site. Findings are reported.

2. AUSTRALIAN INTERNET USAGES

The intensity of Internet use by Australians is among the highest in the world. Australia ranks sixth in the world in terms of the total number of Internet users, despite its relatively small population (4). At the end of September 2003, total Internet subscribers in Australia numbered over 5.2 million, an increase of 135,000 (3%) from the end of March 2003. This modest increase follows the larger increases recorded in the previous two collections, 11% at the end of March 2003 and 8% at the end of September 2002. Majority of these were in the household market with over 3.9 million subscribers. Every second household has home Internet access. (5,9). Australian businesses earned A\$43 billion (USD24 billion) in online revenue from December 2000 to mid 2002 (5).

Australian firms using e-business are saving between one to five percent of ongoing costs, while 10 percent of the businesses save 15 percent or more. (*Allen Consulting Group, Built for Businesses, Australia's Internet Economy, June 2001*). Online purchases of Web services such as recruiting and banking are becoming increasingly popular. (*NUA Internet Surveys, 26 July 2002*). Australian B2B e-commerce transactions were worth \$6.2 billion in 2001 and likely to rise to \$87 billion by 2006 (*International Data Corporation March 2002 Report "Australia B2B eCommerce forecast, 2001-2006*). At the end of March 2002, the number of registered online banking users almost doubled to reach 5.23 million, up from 2.77 million in 2001. The total number of online transactions for the quarter ending March 2002 was 64.3 million, up 7 percent on quarter 1 of 2001 (*NUA Internet Surveys, 10 July 2002*). Usage of Internet banking has grown up to 36 percent in June 2003 from 15 percent in December 2000. Purchasing goods and services increased by 4 percent in June 2003 compared to 2000.(10).

Australian businesses and IT leaders representing medium and large organisations rate security as either "very important" or "important". IDC Survey reports, Governments, business and individuals are moving forward with a heightened sense of security and the inherent interdependent responsibilities. The study also revealed that investment in IT security is increasingly

being driven by strategic initiatives and less in response to security breaches to the organisation. Key drivers behind IT security investment are (i) Increased Internet and Intranet usage and (ii) Mobile computing

3. SECURITY STANDARDS AND POLICIES

The importance of security standardization and certification for e-commerce industry remains an important issue for businesses intending to conduct e-commerce over the Internet. Implementation of security systems must address threats on an ongoing basis. Standard security practices like ISO 17799 and AS/NZS/BS-7799 are used to prevent, detect and contain security breaches and obtain compliance certification. These standards cover critical aspects of good organisational security, which contains a comprehensive set of security controls to improve the level of security within any organisation. The standard is based on assuring integrity, availability, and confidentiality of information assets.

Security is not just the technology alone, but also having solid security policies and following ideal security practices. Without clear standard security policies and practices, e-commerce organisations will not be able to gain the trust of customers or business partners. Therefore the security needs will have to be provided as a set of centralized managed services ensuring corporate interface and information protection. As the e-commerce industry expands, the demands on security level also increases. Business requirements for security changes and the control to threats have to be addressed. It is better to utilise an approach based upon best practices, which has some form of track record, to prevent, detect and contain security breaches. Some security standards in practice serve as a single reference point for identifying a range of controls needed for most situations where information systems are used in industry and commerce.

The author experienced some limitations in gathering information relating to security breaches occurred due to security safeguards and inadequate security management. Hence, it is not possible to document specific scenarios and the measures taken to overcome such breaches. Data is collected through interviews, business websites, publications/white papers related to the issues concerning security and privacy. Though organisations are not willing to reveal previous security breaches, one could see the number of changes introduced in the security issues in the last few years.

4. EVALUATION METHODOLOGY

The triangular relationship between customer, e-business and data security is very simple and clear. In order to safely conduct business and exchange information through the use of technology, we still must be certain that it is as safe as exchanging information

face-to-face. As data loss and breach of privacy are serious concerns in e-trade, it is necessary for the customers to apply a few criteria to evaluate sources to reduce the risk of losing valuable information. In this paper, the author identified specific areas in e-business and evaluated the contents of these web sites. The study also helped in getting a broader understanding of data security adopted by businesses in Australia.

4.1 Methodology

- § Defining the task
- § Selecting known businesses within the industries chosen
- § Selecting a set of criteria for evaluation
- § Evaluating the sources using the criteria
- § Writing a conclusive statement using the information gathered

4.2 Defining the Task

To locate information, two types of industries actively operating in e-business, have been identified. They are

1. *Banking industry* - for its necessity of state-of-the-art security in all online transactions/transmissions

2. *Retail industry* – as it tends to operate on a smaller scale, and generally assumed to have less security.

4.3 Selecting businesses within the industries chosen

Two major banks in Australia, namely, *Westpac Banking Corporation* and *Commonwealth Bank of Australia* were chosen for the Banking Industry. For Retail Industry, *Woolworths Home Shop*, a major super market and *Roses Only*, a leading online florist, were chosen.

4.4 Selecting criteria to evaluate potential sources

The Three Basic criteria chosen to evaluate information security in e-commerce industry are:

- Confidentiality
- Integrity
- Availability

Confidentiality is ensuring that information is accessible only to those who are authorized to have access. Integrity is ensuring accurate, reliable and complete information protected from unauthorized alteration; and Availability is assuring information and associated assets accessible only to authorized users.

The above three criteria further elaborated into:

Criteria	Explanation
Relevance	How is this relevant to data security in eBusiness
Authority	Where is this information coming from?
Currency	When was this information published/updated? What has changed since this was published?
Scope	Does the information cover all aspects of Security issues? To what detail are aspects of this covered?
Accuracy	How accurate is the information provided? Can it be verified?

5. SOURCE EVALUATION

5.1 Banking Industry

Criteria	Westpac	Commonwealth Bank of Australia
Relevance	The policy states, “We are committed to protecting and maintaining the privacy, accuracy and security of your personal and financial information”. Uses up-to-date and secure technology methods to protect information online using multiple firewalls, secure passwords and sign in processes. Personal information is passed through secure server using encryption technology. The site provides a guarantee “when you use Internet Banking, you can be confident that we employ the highest level of security to protect your accounts and personal information”. To confirm that the data is encrypted, a symbol of a lock appears at the bottom.	The Bank takes care of the protection of customers’ information and transaction instructions seriously. As part of maintaining the highest level of customer service, the Commonwealth Bank has implemented system security controls, encryption of sensitive data and regular security reviews as well as a five minute time-out on Net Bank to avoid anyone else accessing your banking details on your computer. Net Bank assures the customers of safety, security, privacy and protection. “We guarantee it!” This means that the customers will not be liable for any unauthorised transaction as a result of using NetBank . It uses 128-bit encryption of all customer data. Customers use Individual Client Numbers and Passwords. The Commonwealth Bank engaged Admiral Computing (Australia) to independently assess the security measures the Commonwealth Bank developed.
Authority	It is found in its official Website: www.westpac.com.au	Issued by: Commonwealth Bank of Australia ABN 48 123 123 124, www.commbank.com.au
Currency	Current. The Website is committed to bring	Current. Electronic Banking Product Disclosure

	customers up to date on the latest security features of online banking.	Statement Date: 5 February 2004
Scope	Describes security features extensively. The security section gives detail information on how to “protect yourself while banking online.”. It also provides latest news on fraud issues, online security and safeguarding customer computer.	<i>Security and Privacy Statement</i> explains how customers’ personal information will be treated as they access and interact with the Web Site. Information covers aspects of security, information collection and handling, how a breach could happen and how to prevent it.
Accuracy	“Ask Westpac” provides answers to questions about its products or services.	The Support Centre answers FAQs. Customers are further supported by NetBank Help Desk on 13 2828 .

5.2 Retail Industry

Criteria	Woolworths	Roses only
Relevance	First level of security is customers login into Woolworths HomeShop with a username and password. Assures that the username and password are examined in a secure session and cannot be accessed by anyone else. This is the second level of security. While browsing around the shopping aisles, no secure data is being passed around and while checking out of the supermarket, another secure session allowing the purchase process to happen safely and privately. It uses the Secure Sockets Layer (SSL). This ensures that the information sent between customer browser and the Webserver cannot be read by others. Secure sessions allow determining the correct products and prices and when the customers purchase products.	<i>Rosesonly.com.au</i> is a secured site using a VeriSign Digital Certificate. This ensures that all information will be encrypted. If any other user intercepts the communication, they will only be able to view an encrypted form. This makes it almost impossible to be intercepted by an unauthorised party. It explains that when the customers order on <i>rosesonly.com</i> , Firstly, a security notice will appear. Also, a secure icon will appear on the browser. For Microsoft Internet Explorer, a lock icon  and Netscape it is a key icon.  The Secure site can also be identified by the address. 'https://'. The industry standard for encryption technology, SSL is used which is compatible with Internet Explorer, Netscape and most other browsers.
Authority	Woolworths Limited http://www.homeshop.com.au/info/SecurityTechFAQ.asp	http://www.rosesonly.com.au/Security.asp#sec8
Currency	Last Updated: Monday, 20 September 2004	Current 2004 Roses Only
Scope	Covers most aspects of Security issues. Explains in detail the use of cookies and Java script.	Security and Privacy section highlights information about Collection and Use of information, cookies, privacy in emails, security and browsers.
Accuracy	For Questions or problems regarding Woolworths HomeShop, call customer representatives on 1300 666 377	www.rosesonly.com.au is a Verisign Secure site. Its validity period is from 06 Nov 03 –05 Nov 04 (checked on 22 Sep 2004)

6. CONCLUSION

The degree of satisfaction through online should be equal to the service they get in an offline environment. This paper addresses a need for a framework to better understand customers’ online trust. The model reflects consumers’ expectations. Businesses can develop e-business strategies to attract more customers through trust in their web sites. As a new form of conducting business over the Internet, e business involves uncertainty and risk than traditional way of doing business. It is obvious that all the businesses mentioned in this paper continue to maintain their eBusiness practices with superior quality, accuracy and currency. It can be concluded that each of the

four businesses place major emphasis on their privacy and security policies. Each use the strongest form of SSL encryption (128-bit), which provides a very high level of protection against unauthorised access and prevent unauthorised persons from reading information customers send to the business, while it is in transit over the Internet. Ensuring security gives confidence and assurance to the customer for a safe transaction. In answering the question *how can we best evaluate e-business information for security purposes?* The above four examples from two major industries would provide some insight in finding information that could be trusted from eBusiness web sites. All these websites identify data security measures implemented and address issues of privacy. All the sites

offer considerable amount of customer advice and information on their methods of protection. Some of the sites mention possible instances of where data security could be breached and how to identify them (i.e. Fraud, hoax emails). The banking industries have shown they have highlighted strongly on the issues of security and protecting of customers privacy. This is mainly due to the fact they are dealing with valuable assets and are needed to reassure customers, especially potential clients who will be investing through their banking system.

REFERENCES

[1].<http://www.idc.com.au/resources/press/software/20>

021105_Security.htm

[2].www.anz.com.au

[3].<http://www.telstra.com.au/ordering/tour/security.htm>

[4].http://www.ozweb.biz/home_internetusage.html

[5].<http://www.nua.com/surveys/index>

[6].<http://www1.ap.dell.com/content/topics/topic.aspx/ap/policy/en/privacy?>

[7].<http://telstra.com/hosting/products/security/>

[8].<http://www.stgeorge.com.au/>

[9].<http://www.abs.gov.au>

[10].www2.dcita.gov.au/publications/2004/01/scope/e-service

Internet Security and Digital Certificates: How Much Do You Know About Them?

Nena Lim

UQ Business School, The University of Queensland, Brisbane, QLD 4072, Australia
n.lim@business.uq.edu.au

ABSTRACT

The objective of this paper is to investigate to what extent the general public knows about Internet security and digital certificates as well as how likely they would use digital certificates if the certificates are free and ready for them to use. The Hong Kong SAR government introduced a new multi-application smart identity card to its residents in August 2003. As the government wants to encourage its residents to adopt electronic commerce, residents can choose to embed a free digital certificate in their smart identity cards. A survey was conducted in Hong Kong SAR in January 2004 and one hundred and forty questionnaires were returned. Results show that the respondents' self-reported knowledge of several key concepts related to Internet security were inconsistent. Moreover, respondents did not seem to have much understanding of the usefulness of the digital certificates nor did they appear to be too keen to use the digital certificates for electronic transactions.

Keywords: digital certificate, electronic certificate, PKI, Hong Kong SAR smart identity card

1. INTRODUCTION

Internet security has been repeatedly shown to be a major worry of consumers and an obstacle for the success of electronic commerce [1-3]. One of the solutions researchers and businesses have found for this problem is the adoption of digital certificates (also known as electronic certificates) for Internet communications and transactions. A digital certificate establishes an individual or an organization's identity in the cyber world. Issued by a certification authority (CA), each digital certificate includes information such as the name of the certificate holder/organization, a unique serial number, the expiration date of the certificate, a copy of the certificate holder/organization's public key, and a digital signature of the certification authority. The mechanism of digital certificates relies on the public key infrastructure (PKI) to achieve on-line authentication. Under the PKI, individuals with no prior relationship can achieve confidentiality and authentication when they swap information over the Internet or conduct electronic commerce transactions [1].

Although PKI has already existed for several decades, many consumers still hesitate to get involved in electronic commerce transactions. Why? One possible explanation is that although public key encryption algorithms have progressed quickly since 1970s [4], consumers do not feel any safer because they do not know how PKI works nor do they understand the protection digital certificates can offer. The objective of this paper is to investigate to what extent the general public knows about Internet security and digital certificates as well as how likely they would use digital certificates if the certificates are free and ready for them to use.

2. HONG KONG SAR SMART IDENTITY CARD

The Hong Kong Special Administration Region (SAR) has a population of 6.8 million in mid-2003 [5]. Sixty-eight percent of households in Hong Kong SAR have personal computers and nearly all these households (89%) have Internet access [6]. The Hong Kong SAR government introduced a new multi-application smart identity card to its residents in August 2003. In addition to immigration purposes such as facilitating automated immigration clearance at border control points, a Hong Kong SAR smart identity card also includes several non-immigration applications.

Residents can now use the digital certificates embedded in their identity cards to update or check personal data kept by the government. With the embedded digital certificates, residents can also encrypt email messages, conduct on-line share trading and on-line betting etc. [7]. Residents can use the digital certificate free for one year because the Hong Kong SAR government wants to encourage people to adopt electronic commerce [8]. Despite the good intentions of the government, do the Hong Kong SAR residents understand what digital certificates are? Are they ready to trust and hence utilize the digital certificates for electronic commerce transactions? To what extent will the goal of Hong Kong SAR government of creating a critical mass for electronic commerce be achieved?

By the beginning of August 2004, over 420,000 smart identity card holders have opted for the free digital certificate [9]. This figure equals to only six percent of the entire population. Assuming more people opt for the digital certificate after the promotion of some renowned celebrities, the key issue one should remember is: a high Internet connectivity rate or even ownership rate of digital certificate does not necessarily lead to individuals' acceptance of electronic commerce.

Citizens in Finland do not consider it worthwhile to pay 10 Euros for a digital certificate [10]. Similarly, a low acceptance rate of digital certificate was reported in Germany. Less than one percent of students in a German university applied for a smart card which has a digital certificate embedded in it. The cards and card readers were heavily subsidized by the university. Yet even computer science students did not show too much interest in the cards. Those who got the cards only used them infrequently [10].

3. METHODOLOGY

To investigate their understanding of Internet security and digital certificates as well as their readiness in using the free digital certificates, a survey of Hong Kong residents was conducted in January 2004. One hundred and fifty questionnaires were distributed to associate degree students who have enrolled in an advanced accounting course at the City University of Hong Kong. One hundred and forty questionnaires were returned. Among the respondents, 72 percent were female and 27 percent were male. They aged from 20 to 23. The average age was 20.8. Most respondents had five to eight years experience of using computer and five to six years experience of using the Internet. Ninety-nine percent had computers at home.

4. RESULTS

In a seven-point Likert scale, the average self-reported knowledge of Internet security in general was 3.7 and only one respondent indicated he/she had no knowledge of Internet security. Yet interestingly respondents' self-reported knowledge of PKI was much lower with an average score of only 2.3. Fifty-four respondents indicated that they knew nothing about PKI. Similar results exist for respondents' self-reported knowledge of certification authorities and digital certificates. The respondents' average knowledge of certification authorities and digital certificates were 2.35 and 3.3 respectively. The relatively higher self-rating of digital certificate is unsurprising because ten percent of the respondents already owned valid digital certificates. These people must have obtained the certificates on their own because according to the government schedule it was not their turns to apply for the new identity cards yet. The results raise an interesting question on the inconsistency of respondents' self-reported knowledge. Did people really know about Internet security or did they over-estimate their understanding?

There is no significant difference between the "have" and "have not" groups of digital certificates in all aspects except their knowledge of digital certificates. Out of the 125 respondents who did not have a valid digital certificate, two-thirds did not know what applications digital certificates could provide. After being told the applications they could use, these respondents were not keen to use those applications. For

example, while 119 respondents use electronic mails regularly, only 20 percent indicated that they might use the digital certificates to encrypt their emails. Furthermore, two-thirds of these respondents were not aware that they would need to have a smart-card reader to use the free digital certificate embedded in their smart identity card for on-line applications. Once they were aware of the equipment requirement, 95 percent indicated they were either unsure or unlikely to purchase a smart card reader. Such results are unsurprising as smart card readers have been shown to be a major obstacle of smart card adoption [10]. Although 40 percent of respondents indicated they would opt to include the free digital certificates into their identity cards, 93 percent of them would not pay the annual fee themselves once the certificates expire after one year. Results also indicate that those who were willing to buy smart card readers are more likely to pay annual fees ($r = 0.425$).

5. CONCLUSIONS

In summary, results of this study show that the self-reported knowledge of respondents of several key concepts related to Internet security were inconsistent. It is doubtful how much they really know. Moreover, respondents did not seem to have much understanding of the usefulness of the digital certificates embedded in their new smart identity cards and they did not appear to be too keen to use the digital certificates for electronic transactions. To achieve its objective of promoting electronic commerce transactions, we recommend the Hong Kong SAR government to consider further educating its residents on Internet security because results of this study suggest that the general public in Hong Kong is unlikely to make good use of the digital certificates. While some forms of public education such as exhibitions were conducted in the past, results suggested that more should be done. For example, promotions of useful applications such as on-line submission of tax returns can be used to attract usage of digital certificates. Moreover, more advertising and education are required to familiarize residents with the benefits of digital certificates. One way to educate the general public is to start from students, from primary schools to tertiary institutions, so that the next generation will be more ready to face the electronic age.

ACKNOWLEDGEMENT

The author would like to thank Carol Kwong for her assistance in data collection.

REFERENCES

- [1] Greenstein, M., and Vasarhelyi, M., *Electronic Commerce: Security, Risk Management, and Control*, McGraw-Hill, Irwin, second edition, 2002.

- [2] Lim, N., "Consumers' perceived risk: Sources versus Consequences", *Electronic Commerce Research and Applications*, Vol. 2, No. 3, pp216-228, 2003.
- [3] Vijayasathy, L. R., "Predicting consumer intentions to use on-line shopping: The case for an augmented technology acceptance model", *Information & Management*, Vol. 41, pp747-762, 2004.
- [4] Schneider, G. P., *Electronic Commerce*, Course Technology, Thompson, 2003.
- [5] The Census and Statistics Department, "Hong Kong populatin projections 2004-2033", *Hong Kong Monthly Digest of Statistics*, Hong Kong SAR, 2004.
- [6] The Census and Statistics Department, *Household survey on information technology usage and penetration*, Hong Kong SAR, 2004.
- [7] Hongkong Post, "Using e-cert on smart ID card" <http://www.hongkongpost.gov.hk/activity/smartid/carduse/index.html>.
- [8] Mak, S. "Enhancing information security with digital certificate and public key infrastructure", *Presentation at Managers Forum-IT Series*, 2003.
- [9] Hongkong Post, *Press Release*, <http://www.hongkongpost.gov.hk/news/press/39.html>, Hong Kong Post, 2004.
- [10] Balaban, D. "Digital signature cards: For professionals only?", *Card Technology*, Vol. 8, No. 3, pp28-30, 32, 34, 2003.

Secure Mobile Agents in Electronic Commerce by Using Undetachable Signatures from Pairings

Yang Shi, Xiaoping Wang, Liming Cao, Jianxin Ren

Department of Computer Science and Technology, Tongji University, Shanghai 200092, China
cnshtyang@yahoo.com.cn

ABSTRACT

It is expected that mobile agents technology will bring significant benefits to electronic commerce. But security issues, especially threats from malicious hosts, become a great obstacle of widespread deployment of applications in electronic commerce based on mobile agents technology. Undetachable digital signature is a category of digital signatures to secure mobile agents against malicious hosts. An undetachable signature scheme by using encrypted functions from bilinear pairings was proposed in this paper. The security of this scheme is based on the computational intractability of discrete logarithm problem and computational Diffie-Hellman problem on gap Diffie-Hellman group. Furthermore, the scheme satisfies all the requirements of a strong non-designated proxy signature i.e. verifiability, strong unforgeability, strong identifiability, strong undeniability and prevention of misuse. An undetachable threshold signature scheme that enables the customer to provide mobile agents with 'shares' of the undetachable signature function is also provided. It is able to provide more reliability than classical undetachable signatures.

Keywords: electronic commerce, mobile agents, undetachable digital signatures, bilinear pairings

1. INTRODUCTION

Mobile agents technology are attracting a great deal of interest from both industry and academia since middle of 1990's. Compared with traditional computing models, e.g. client/server, mobile agents technology has following advantages[1][2][3]:

- ò Autonomous mobile agents strive to achieve a given goal without permanent observation by its owner. As a matter of consequence, the user is free to take care of other tasks, saving time in the process.
- ò If a host is being shut down, all mobile agents executing on that machine are warned and given time to dispatch and continue their operation on another host in the network.
- ò Users may dispatch mobile agents over a temporary network connection to a target network. After dispatching, the temporary network link may be brought down until a later point in time.

Mobile agents technique brings significant benefits to electronic commerce because of these advantages. But on the other hand, there are also some problems. The most important one is security.

Threats to the security of mobile agents generally fall into four comprehensive classes[4]:

- ò Agent against agent platform
- ò Agent platform against agent
- ò Agent against other agents
- ò Other entities against agent system

Hohl[5] identified the following attacks: spying out code; spying out data; spying out control flow; manipulation of code; manipulation of data;

manipulation of control flow; incorrect execution of code; masquerading of the host; denial of execution; spying out interaction with other agents; manipulation of interaction with other agents; returning wrong results of system calls issued by the agent.

Thus, security issues, especially threats from potentially malicious hosts become a great obstacle of widespread deployment of applications in electronic commerce based on mobile agents technique.

2. PREVIOUS WORKS ABOUT UNDETACHABLE DIGITAL SIGNATURES

2.1 The preliminary idea

Before 1998, many researchers believed that, on malicious hosts, mobile agents were impossible to prevent tampering unless trusted and tamper-resistant hardware is available. Following points[6] are considered by them:

- ò Cleartext data can be read and changed.
- ò Cleartext programs can be manipulated.
- ò Cleartext messages can be faked.

But Sander and Tschudin[7] pointed out that this belief is incorrect because mobile agents do not have to be executed in cleartext form. They proposed the idea of undetachable digital signatures that allows a mobile agent to effectively produce a digital signature inside a remote and possibly malicious host without the host being able to deduce the agent's secret or to reuse the signature routine for arbitrary documents. Here is a brief introduction of the idea.

Let Sig be a rational function used by C (a customer) to produce the digital signature $Sig(m)$ of an arbitrary message m . Furthermore suppose the message m is the result of a rational function f applied to some input data x . Finally the verification function Ver that C publishes in order to let others check the validity of the digital signature z is regarded to be a valid signature of m if and only if:

$$z = Sig(m) \tag{1}$$

For letting the customer's mobile agent create "und detachable" signatures, he computes:

$$f_{Signed} = Sig \circ f \tag{2}$$

Then he sends f_{Signed} and f to S (a shop) with his mobile agent. S evaluates:

$$m = f(x) \tag{3}$$

$$z = f_{Signed}(x) \tag{4}$$

Though the signature function, Sig , is not known by others, every one can verify the validity of a message m by testing:

$$Ver(z) = m \tag{5}$$

2.2 The first implementation

Although Sander and Tschudin tried to give an outline of und detachable digital signatures by using birational functions based on Shamir's work[8]. Unfortunately no secure und detachable digital signatures scheme has been proposed until 2000. In 2000 Kotzanikolaou, Burmester and Chrissikopoulos presented an RSA implementation[9] of und detachable digital signatures. But this scheme does not provide server's non-repudiation because it does not contain server's signature[10].

2.3 Strong proxy signatures and other implementations

In 2001 Lee, Kim and Kim provided an RSA based construction of und detachable digital signatures called "Strong Non-designated Proxy Signature" [10]. Their scheme enhanced [9] and often be acronymized as "LKK-SPS" scheme. A scheme of und detachable threshold signature[13] was proposed by Borselius, Mitchell and Wilson in the same year. In 2002, a strong proxy signature scheme with proxy signer privacy protection[11] was given and a pragmatic alternative to und detachable signatures[12] was also proposed.

3. BILINEAR PAIRINGS AND SIGNATURE SCHEMES BASED ON THEM

3.1 Mathematical preliminaries of Bilinear Pairings

Let G_1 be a cyclic group generated by P , whose

order is a prime q , and G_2 be a cyclic multiplicative group of the same order q : The discrete logarithm problems in both G_1 and G_2 are hard. Let $e : G_1 \times G_1 \rightarrow G_2$ be a pairing satisfies the following conditions:

ò Bilinear: (6) and (7) or (8)

$$e(P_1 + P_2, Q) = e(P_1, Q)e(P_2, Q) \tag{6}$$

$$e(P, Q_1 + Q_2) = e(P, Q_1)e(P, Q_2) \tag{7}$$

$$e(aP, bQ) = e(P, Q)^{ab} \tag{8}$$

ò Non-degenerate: There exists $P \in G_1$ and $Q \in G_1$ subject to (9).

$$e(P, Q) \neq 1 \tag{9}$$

ò Computability: There is an efficient algorithm to compute $e(P, Q)$ for all $\{P, Q\} \subseteq G_1$

We note that the Weil and Tate pairings associated with supersingular elliptic curves or abelian varieties can be modified to create such bilinear maps. Suppose that G is an additive group. Four mathematical problems is defined as follow [14][15].

ò Discrete Logarithm Problem (DLP):

Given two group elements P and Q , and an integer n , such that (10) is satisfied whenever such an integer exists.

$$Q = nP \tag{10}$$

ò Decision Diffie-Hellman Problem (DDHP):

For $\{a, b, c\} \subseteq Z_q^*$, given P, aP, bP, cP decide whether:

$$c \equiv ab \pmod{q} \tag{11}$$

ò Computational Diffie-Hellman Problem (CDHP):

$$\text{For } \{a, b\} \subseteq Z_q^* \quad P, aP, bP$$

compute abP

ò Gap Diffie-Hellman Problem (GDHP):

A class of problems where DDHP is easy while CDHP is hard.

We assume through this paper that CDHP and DLP are intractable, which means there is no polynomial time algorithm to solve CDHP or DLP with non-negligible probability. When the DDHP is easy but the CDHP is hard on the group G_1 , G_1 is called a Gap Diffie-Hellman (GDH) group. Such groups can be found on supersingular elliptic curves or hyperelliptic curves over finite fields, and the bilinear pairings can be derived from the Weil or Tate pairing $e : G_1 \times G_1 \rightarrow G_2$. Our schemes of this paper can be built on any GDH group. More mathematical background can be found in [14][16][17][18]. Now, some system parameters should be defined for Sec. 4 as following: Let P be a generator of G_1 , the bilinear

paring is $e : G_1 \times G_1 \rightarrow G_2$. Moreover, to hash functions are given here: $H_1 : \{0,1\}^* \rightarrow Z_q$ and $H_2 : \{0,1\}^* \rightarrow G_1$. The implementation of these hash functions can be referred to works such as [25]. But the choice of conventional hash functions should be very carefully because many hash functions were cracked recently [26].

3.2 A Survey of Signature Schemes Based on Bilinear Pairings

In recent years, variety schemes of digital signatures have been proposed. For threshold signatures, Boldyreva[15] proposed a robust proactive threshold signature scheme, a multi-signature scheme and a blind signature scheme which work in any Gap Diffie Hellman group in 2002. Then Vo, Zhang and Kim[19] presented a new threshold blind digital signature based on pairings without the third party. For Multisignatures, Lin, Wu and Zhang[20] proposed a new structured multi-signature scheme from the Gap Diffie-Hellman Group that considers the signing order among co-signers. Boneh, Boyen and Shacham[21] constructed a short group signature scheme based on the Strong Diffie-Hellman assumption. Nguyen[22] proposed a group signature scheme with constant-size parameters that does not require any trapdoor secret, thereby, allows sharing of public parameters among organizations. As to Blind signature, Chow et al.[23] gave an unlinkable partially blind signature scheme and an ID-based unlinkable partially blind signature scheme. Furthermore, Zhang et al.[24] addressed that it was easy to design proxy signature and proxy blind signature from the conventional ID-based signature schemes using bilinear pairings and gave some concrete schemes based on existed ID-based signature schemes.

4. UNDETACHABLE SIGNATURES BASED ON BILINEAR PAIRINGS

4.1 Settings

Assume that all participants have the common system parameters: $(G_1, G_2, e, q, P, H_1, H_2)$.

Settings about the customer :

- ò Let C be an identifier for the Customer.
 - ò Let s_C be the signature key of C .
 - ò Let PU_C be the verification key of C where:
- $$PU_C = s_C P \quad (12)$$
- ò Let req_C be the constraints of the Customer.
 - ò Let T_C be the timestamp of req_C

Settings about shops :

- ò Let S be an identifier for a shop.

- ò Let s_S be the signature key of S .
 - ò Let PU_S be the verification key of S where:
- $$PU_S = s_S P \quad (13)$$
- ò Let bid_S be the bid information of S .
 - ò Let T_S be the timestamp of bid_S .
 - ò Let $\varphi : G_1 \rightarrow \{0,1\}^*$ be a function mapping a point of G_1 into a binary string.

4.2 Security scheme

The Customer does following operations :

- ò Computes:
- $$H_C = H_2(C \parallel req_C \parallel T_C) \quad (14)$$

- ò Computes:
- $$K_C = s_C \cdot H_C \quad (15)$$

- ò Gives $(C \parallel req_C \parallel T_C)$ to the Agent.

- ò Gives to the Agent as part of its executable code the undetachable signature function pair:
- $$f(\cdot) = (\cdot)H_C \quad (16)$$

and

$$f_{Signed}(\cdot) = (\cdot)K_C \quad (17)$$

Suppose :

$$Sig(x) = s(\cdot) \quad (18)$$

It turns out the following proposition:

Proposition 1. The functions above construct an undetachable digital signature scheme.

Proof.

$$\begin{aligned} & (Sig \circ f)(\cdot) \\ &= Sig(f(\cdot)) \\ &= Sig((\cdot)H_C) \\ &= s_C((\cdot)H_C) \\ &= (\cdot)s_C H_C \\ &= (\cdot)K_C \\ &= f_{Signed}(\cdot) \end{aligned} \quad (19)$$

After the Agent migrated to the host of S , the validity of the mobile agent should be verified first by checking:

$$e(K_C, P) \stackrel{?}{=} e(H_C, PU_C) \quad (20)$$

If it is a valid signature of C , then S does following operations :

- ò Computes:
- $$H_S = H_2(C \parallel req_C \parallel T_C \parallel S \parallel bid_S \parallel T_S) \quad (21)$$

- ò Computes:
- $$Y = s_S H_S \quad (22)$$

- ò Computes: $x = H_1(\varphi(Y))$ (23)
- ò Computes: $z = f_{Signed}(x)$ (24)
- ò Gives $(x, z, Y, H_S, S, bid_S, T_S)$ to the mobile agent.

After the mobile agent goes back to C . The correctness of (21) and (23) should be verified first. Then the verification should be performed by formula (25):

$$e(Y)P(=e H_S, PU_S) \quad (25)$$

4.3 Security Analysis

The following proposition can be obtained because G_1 is a GDH group.

Proposition 2. A transaction is valid if and only if (20), (21), (23) and (25) are true.

Proof. If the mobile Agent is not detached before it migrated to the host of S , then:

$$\begin{aligned} & e(K_C, P) \\ &= e(s_C H_C, P) \\ &= e(H_C, P)^{s_C} \\ &= e(H_C, s_C P) \\ &= e(H_C, PU_C) \end{aligned} \quad (26)$$

If an opponent Oscar want to modify C 's bid information when C 's agent mobiles to his host, he has to construct a new undetachable digital signature pair (H_C, K_C) of the Customer which will include modified constraints req_C of the Customer. But this needs to solve the computational difficult problems mentions in section 3.1.

Furthermore, if a transaction is valid:

$$\begin{aligned} & e(Y)P \\ &= e(s_S H_S, P) \\ &= e(H_S, P)^{s_S} \\ &= e(H_S, s_S P) \\ &= e(H_S, PU_S) \end{aligned} \quad (27)$$

Similar to equation (20), the security of (23) and (25) also relies upon the difficulty of the problems computational infeasible to solve at present.

5. UNDETACHABLE THRESHOLD SIGNATURES

5.1 Basic Idea

The notion of undetachable threshold signatures was introduced in [27]. An undetachable threshold signature scheme will enable the customer, C , to provide n mobile agents with 'shares' of the signature key (where the shares will be a function of req_C and T_C). For more details about undetachable threshold signature such as their usage and application can be found in [27].

5.2 Security Scheme

Based on undetachable signature scheme proposed in section 4, a variety of secret sharing scheme can be used to construct undetachable threshold signatures by converting s_C into signature shares. An example using Lagrange Polynomial Interpolation is given here:

Suppose f is a polynomial over Z_q with degree $t - 1$ subject to (28).

$$s_C = f(0) \quad (28)$$

Operations in equations (29) to (32) and (38) are performed over Z_q .

Let:

$$s_i = f(i), \quad i=1,2,L, n \quad (29)$$

Then any one has t shares can obtain s_C by calculating (30)

$$s_C = \sum_{j=1}^t \left(s_{i_j} \frac{\prod_{h=1, h \neq j}^t i_h}{\prod_{h=1, h \neq j}^t (i_h - i_j)} \right) \quad (30)$$

To simplify the description we suppose the t shares are s_1, L, s_t , thus:

$$s_C = \sum_{j=1}^t (s_j) \quad (31)$$

Where:

$$l_j = \frac{\prod_{h=1}^t h}{\prod_{h=1, h \neq j}^t (h - j)} \quad (32)$$

Then C does following extra operations

- ò Computes: $PC_i = s_i P, \quad i=1,L, n$ (33)

and (34) instead of (14).

$$H_C = H_2(C \parallel req_C \parallel T_C \parallel PC_1 \parallel L \parallel PC_n) \quad (34)$$

- ò Computes: $K_i = s_i H_C, \quad i=1,L, n$ (35)

ò Gives $PC_1 \parallel L \parallel PC_n$ to these n agents.

Finally, C gives (36) to the i-th agent as part of its executable code instead of (17):

$$f_{Signed_i}(\cdot) = (\cdot)K_i \quad (36)$$

After the i-th Agent migrated to the host of S , the validity of the mobile agent should be verified first the by checking:

$$e(K_i, P) \stackrel{?}{=} e(H_i, PC) \quad (37)$$

It is clear that a shop has t or more than t shares can reconstruct f_{Signed} by (38). But anyone cannot reconstruct f_{Signed} from less than t shares.

$$f_{Signed}(\cdot) = \sum_{i=1}^t [v_i f_{Signed_i}(\cdot)] \quad (38)$$

After a shop has reconstruct f_{Signed} successfully, other operations are similar to undetachable signatures proposed in section 4. So redundant words are omitted.

6. CONCLUSION

In this paper, we have presented a novel implementation of undetachable digital signatures and a correspondent security scheme. Compared to [9][10], our scheme uses a different cryptosystem to construct undetachable signatures. This implementation of undetachable digital signatures is based on non-interactive CEF (Computing with Encrypted Functions) from bilinear pairings to protect the original signature function Sig by encrypting it with a function f to obtain the encrypted function f_{Signed} defined as the composition of Sig and f . Furthermore, the scheme satisfies all the five requirements of a strong non-designated proxy signature proposed by Lee et al. in [10] as follow: First, verifiability: A proxy signer can create a valid proxy signature for the original signer. But the original signer and any third party cannot create a valid proxy signature with the name of proxy signer, Second, strong unforgeability: Anyone can determine the identity of the corresponding proxy signer form a proxy signature. Third, strong identifiability: Once a proxy signer creates a valid proxy signature on behalf of an original signer, the proxy signer cannot repudiate his signature creation against anyone. Fourth, strong undeniability: It should be confident that proxy key pair cannot be used for other purpose. In the case of misuse, the responsibility of proxy signer should be determined explicitly. Finally, preventions of misuse: From a proxy signature a verifier can be convinced of the original signer's agreement on the signed message.

As to the undetachable threshold signature scheme proposed in the paper, it has following features: First, each agent can use their share to sign a message M , e.g. a piece of bid information, of their choice to obtain

a 'signature share'. Second, The 'correctness' of a signature share can be verified independently of any other signature shares. Third, any shop, when equipped with t different signature shares restricted with the same request of C for the same message M , can construct a signature on M which will be verifiable by any party with a trusted copy of public key of C , and which will also enable the corresponding req_C and T_C to be verified. Finally, knowledge of less than t different signature shares for the same message M cannot be used to construct a valid signature one the message M , knowledge of any number of different signature shares for messages other than the message M will not enable the construction of a valid signature on M and knowledge of any number of different signature shares for request other than req_C or with a different time stamp will not enable the construction of a valid signature with associated req_C and T_C . So our scheme satisfied all the requirements of undetachable threshold signatures defined in [27] and provide more reliability than classical undetachable signatures.

ACKNOWLEDGEMENT

This work is supported by the National Science Foundation of China under grant 70171061.

REFERENCES

- [1] Puliafito, A., Riccobene, S., Scarpa, M., "An analytical comparison of the client-server, remote evaluation and mobile agents paradigms", *proceeding of First International Symposium on Agent Systems and Applications*, pp278 -292, 1999.
- [2] Busch, C., Roth, V., Meister, R., "Perspectives on Electronic Commerce with Mobile Agents", *proceeding of the XI Amaldi Conference On Problems of Global Security*, pp1-13, 1998.
- [3] Lange, D., B., Oshima., M., "Seven good reasons for mobile agents", *Communications of the ACM*, Vol.42, No.3, pp88-89, 1998.
- [4] Jansen, W., "Countermeasures for mobile agent security", *Computer Communications*, No.13, pp1667-1676, 2000.
- [5] Hohl, F., "Time Limited Blackbox Security: Protecting Mobile Agents from Malicious Hosts", *Lecture Notes in Computer Science*, vol.1419, pp92-113, 1998.
- [6] Chess, D., Grosf, B., Harrison C., et al. "Itinerant agents for mobile computing", Technical Report RC 20010, IBM, March 1995.
- [7] Sander, T., Tschudin, C., "Protecting Mobile Agents Against Malicious Hosts", *Lecture Notes in Computer Science*, vol.1419, pp44-60, 1998.

- [8] Shamir, A., "Efficient signature schemes based on birational permutations", *proceeding of CRYPTO'93*, vol.773, pp1-12, 1993.
- [9] Kotzanikolaous, P., Burmester, M., Chrissikopoulos, V., "Secure Transactions with Mobile Agents in Hostile Environments", *proceeding of ACISP 2000*, pp289-297, 2000.
- [10] Lee, B., Kim, H., Kim, K., "Secure mobile agent using strong non-designated proxy signature". *proceeding of ACISP 2001*, pp474-486, 2001.
- [11] Shum, K., Wei, V., "A strong proxy signature scheme with proxy signer privacy protection", *proceeding of Eleventh IEEE International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*, IEEE, 2002.
- [12] Borselius, N., Mitchell, C., Wilson, A., "A pragmatic alternative to undetachable signatures", *ACM SIGOPS Operating Systems Review*, vol.36, No.2, pp6-11, 2002.
- [13] Borselius, N., Mitchell, J., Wilson, A., "Undetachable threshold signatures", *proceeding of the 8th IMA International Conference*, Cirencester, UK, pp239--244, 2001.
- [14] Boneh, D., Franklin, M., "Identity-based encryption from the Weil pairings", *proceeding of Advances in Cryptology-Crypto*, pp.213-229, 2001.
- [15] Boldyreva, A., "Efficient threshold signature, multisignature and blind signature schemes based on the Gap-Diffie-Hellman-group signature scheme", *proceeding of Practice and Theory in Public Key Cryptography -- PKC'2003*, Lecture Notes on Computer Science, Vol. 2567, Springer-Verlag, pp31-46, 2003.
- [16] Frey, G., Muller, M., Ruck, H.G., "The Tate Pairing and the Discrete Logarithm Applied to Elliptic Curve Cryptosystems", *IEEE Transactions on Information Theory*, vol.45, No.5, pp1717-1719, 1999.
- [17] Frey, G., Ruck, H.G., "A remark concerning the m -divisibility and the discrete logarithm in the divisor class group of curves", *Mathematics of Computation*, vol.62, No.206, pp865-874, 1994.
- [18] Silverman, J.H., *The Arithmetic of Elliptic Curves*, Springer-Verlag, New York, 1986.
- [19] Vo, D. L., Zhang, F., Kim, K., "A New Threshold Blind Signature Scheme from Pairings", *2003 Symposium on Cryptography and Information Security (SCIS2003)*, vol.1, pp233-238, 2003.
- [20] Lin, C.Y., Wu, T.-C., Zhang, F., "A Structured Multisignature Scheme from the Gap Diffie-Hellman Group", <http://eprint.iacr.org>, 2004.
- [21] Boneh, D., Boyen, X., Shacham, H., "Short Group Signatures", *proceeding of Advances in Cryptology -- Crypto'2004*.
- [22] Nguyen, L., "A Trapdoor-free and Efficient Group Signature Scheme from Bilinear Pairings", <http://eprint.iacr.org>, 2004.
- [23] Chow, S. S. M., Hui, L. C. K., Yiu, S. M., et al., "Two Improved Partially Blind Signature Schemes from Bilinear Pairings," <http://eprint.iacr.org>, 2004.
- [24] Zhang, F., Safavi-Naini, R., Lin, C. Y., "New Proxy Signature, Proxy Blind Signature and Proxy Ring Signature Schemes from Bilinear Pairing", <http://eprint.iacr.org>, 2004.
- [25] Barreto, P., Kim, H., "Fast Hashing onto Elliptic Curves over Fields of Characteristic 3", <http://eprint.iacr.org>, 2002.
- [26] Wang, X., Feng, D., Lai, X. et al. "Collisions for Hash Functions MD4, MD5, HAVAL-128 and RIPEMD", <http://eprint.iacr.org/2004/>.
- [27] Borselius, N., Mitchell, J., Wilson, A., "Undetachable threshold signatures", *proceedings of the 8th IMA International Conference*, Lecture Notes on Computer Science, Vol. 2260, Springer-Verlag, pp239-244, 2001.

Security Service Model for RFID Enabled Supply Chain

Bo Chen¹, Hongying Gu², Oliver Wang³

¹ Software College, Zhejiang University of Technology, Hangzhou 310032, China

² Computer College, Zhejiang University, Hangzhou 310027, China

³ SAP China Labs, Shanghai 200003 China

¹ cb@zjut.edu.cn, ² guhy@zju.edu.cn, ³ Oliver.wang@sap.com

ABSTRACT

It has been widely recognized that RFID related technologies will greatly improve the visibility, the efficiency and the collaboration of industry supply chain. In this new “product driven” supply chain scenario, manufacturer, supplier, and third party share and coordinate the use of diverse resources in distributed “virtual organizations”. It challenges the security issues, which demand new technical approaches. In collaboration with researchers in Auto-ID Labs China, we have developed a service-oriented framework based on CA and Web Services, which supports inexpensive mediation of product information among rapidly evolving, heterogeneous information sources. Our architecture defines a user-driven security model that allows users to create entities and policy domains within virtual organizations. We emphasize that standard Web Services tools and software provide both stateless and stateful forms of secured communication.

Keywords: security service, RFID, supply chain, PKI

1. INTRODUCTION

Radio frequency identification (RFID) is a powerful new technology that is helping companies to improve visibility into their inventory data and increase accuracy across supply chain networks. When RFID tags are attached to products, boxes, and pallets, the items can be followed automatically as they move, providing a real-time and accurate view of inventory[1]. The technology is so effective that organizations such as Wal-Mart, Metro, and the U.S. Department of Defense now require their suppliers to deliver goods with RFID tags.

The Auto-ID Center in MIT, was launched in late 1999, has developed RFID tags which are cheap enough to be disposable and readers which are agile enough to be deployed globally[2]. According to Auto-ID Center’s researching, the various components of the Auto-ID infrastructure includes (Figure 1):

- ò Hardware – such as RFID tags and readers but also including optical readers and other types of sensors.
- ò Savant – a software platform for interfacing to various types of readers and filtering the data generated when the tags are within range of the readers.
- ò The Electronic Product Code (EPC) – a unique identification number for each object – the minimum information which must be stored on its ID tag.
- ò The Object name service (ONS) – an extension of the domain name service (DNS), which provides a method of converting an EPC into the network address of the corresponding networked database, which in turn holds the product data for that object.
- ò The EPC Information Server (formerly PML Server) – an open standard XML Schema for representing information about products and

communicating this information globally in a unified manner, between parties who otherwise have disparate databases, data storage formats and who use diverse application programs to act upon the product data.

- ò Control systems – to provide feedback into the supply chain or manufacturing process, using the data generated by EPC readings.

EPC Network Architecture-across Enterprises

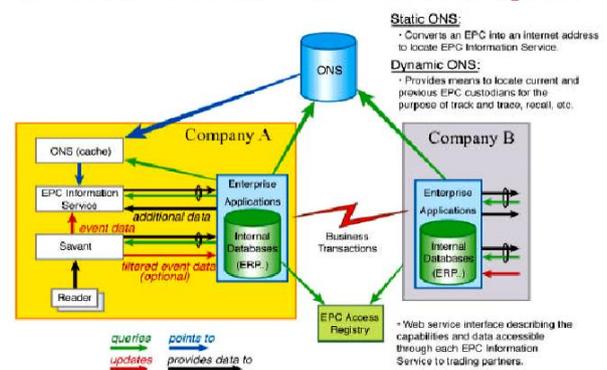


Figure 1 EPC Network Architecture

Source: Autoidcenter.org

Since each unique object will have a unique identifier, in the form of the EPC, each object will become an individual, with its own individual history about how it passed the through supply chain. In this “product driven” supply chain scenario, manufacturer, supplier, and third party (like carrier, distribute center) share and coordinate the use of diverse resources in distributed “virtual organizations”.

Let us consider a simple business scenario. The suppliers and customers (manufacturers) collaborate by sharing information and exchanging plans via the Supplier Managed Inventory (SMI) process. The SMI process is a supplier-driver replenishment and planning process, which is based on min/max stock balance levels, on current demand, in-transits and planned shipments.

With RFID enabled, the application now should support the handling of massive amounts of Auto-ID data through connectivity with readers, tags, and other devices; and integration of high-volume RFID data with back-end business processes. The dynamic and multi-parties nature of these environments introduces challenging security issues that demand new technical approaches. Most of data are sensitive, and every server exposed on internet will be attacked, and so on.

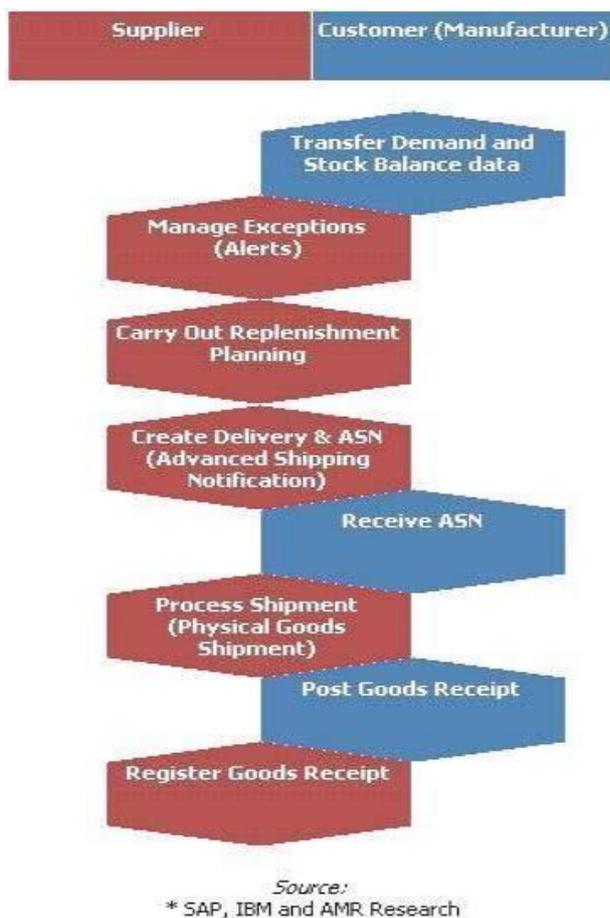


Figure 2 SMI scenario

The remainder of this article is organized as follows. In Section 2, we investigated related work on security web service and certificate validation. Then we propose a design of security system model, emphasizes on security web service and privilege management infrastructure. Finally, we explain function for system and then we conclude this paper with a brief discussion of future work.

2. RELATED WORK

2.1 CA-based Trust Model

One commonly used identity token is provided by an X.509 public key certificate. An X.509 “public key certificate” is defined in ITUT standard [3] as “the public key of a user, together with some other information, rendered unforgeable by encipherment with the private key of the certification authority which issued it.” [Section 3.3.44]

In practice an X.509 certificate as defined by the PKIX X.509 Certificate Profile RFC[] contains a public key, a subject name in the form of a multi-component Distinguished Name [DN] and a validity period and is signed by a trusted third party called a Certification Authority (CA). The X.509 certificates are used with the TLS(SSL) security protocol to make a secure authenticated connection between two parties. X.509 certificates are exchanged between entities, usually in a phase where no security has yet been negotiated (no data integrity or confidentiality). The certificates are first tested for validity by checking the expiration dates, possible revocation, acceptable key usage and if they are signed by trusted CA. If the certificates pass all these checks their public keys are then used to build a challenge handshake to prove that each entity that sent a certificate has the corresponding private key.

There are other public/private key based schemes such as PGP keys[4], SSH keys [5]and the SPKI [6] keys and protocols. The PKIX/X.509 scheme has a smallish set of trusted third parties (CAs) to sign identity certificates that contain a subscriber’s public key. This improves the scaling properties of public key distribution since only the CA’s public key needs to be distributed in an out-of-band secure manner. Another feature of the X.509-TLS infrastructure is that it supports multiple independent CAs. In a supply chain, each company may choose which CAs it will accept for binding DNs and public keys. At this point, most of the current solutions allow for short-term proxy certificates, stored with unencrypted private keys, to which a user has delegated his identity. These certificates are correctly formatted X.509 certificates, except that they are marked as proxy certificates and are signed by an end entity rather than a CA[7].

2.2 XML Web Service Security

The XML security standards define XML vocabularies and processing rules. These standards use legacy cryptographic and security technologies, as well as emerging XML technologies, to provide a flexible, extensible and practical solution toward meeting security requirements. The XML security standards include XML digital signature [8]for integrity and signing solutions, XML encryption [9] for confidentiality, XML key management (XKMS) [10] for public key registration,

location and validation, security assertion markup languages (SAML) [11] for conveying authentication, authorization and attribute assertions, XML access control markup language (XACML) [12] for defining access control rules, and platform for privacy preferences (P3P) [13] for defining privacy policies and preferences. Major use cases include securing web services (WS-Security) [14].

2.3 RBAC Policies in Privilege Management

Role Based Access Controls (RBAC) has generated significant interest in the last decade. The main entities in RBAC are the users, the roles and the permissions. A role can represent a job function, a qualification or expertise. A permission represents the right to access a target in a particular mode. Roles are assigned to users in a many to many relationship, and permissions are granted to roles, again in a many to many relationship. There is thus a level of indirection between a user and his access rights. The benefits of RBAC are widely discussed on lots of papers [15].

Edition 4 of X.509 [3], published in 2001, is the first edition to fully standardize the certificates of a Privilege Management Infrastructure (PMI). X.509 supports RBAC by defining role specification attribute certificates that hold the permissions granted to each role, and role assignment attribute certificates that assign various roles to the users. Each role and name component in X.509 and LDAP [16] is an attribute type, attribute value pair. Thus roles and name components are easily interchangeable. The user is identified by either his/her LDAP Distinguished Name or his public key certificate (issuer and serial number). Role assignment ACs may point to their corresponding role specification AC via the role specification certificate identifier extension.

A policy is defined which states the rules for assigning roles to users, and permissions to roles. The policy can then be used to control the accesses to all the targets within the policy domain [17]. In practice, now we have APIs that read in the XML Policy, parse it, and use it to control access to targets within the policy domain. The API caller, typically an application gateway, passes the authenticated name of the user, and this is used to retrieve the user's attribute certificates (ACs) from the configured directory (In most cases, LDAP server is used). The API caller passes the user's requested action on his chosen target, and again this is checked against the policy. The API returns either granted or denied to the caller.

3. SECURITY MODEL FOR RFID-ENABLED SCM

We now turn to outline the proposed structure of a prototype SSS (SCM Security Server).

Web services technologies allow software components to be defined in terms of access methods, bindings of these methods to specific communication mechanisms, and

mechanisms for discovering relevant services. Now, the Simple Object Access Protocol (SOAP) and the Web Services Description Language (WSDL) are widely used in many applications.

SSS defines standard Web service interfaces and behaviors that add to Web services the concepts of stateful service and secure invocation, as well as other capabilities needed to address RFID-specific requirements. These interfaces and behaviors define what is called a "SSS Service" and allow users to manage the web service's life-cycle, as allowed by policy, and to create sophisticated distributed services.

Our SSS security model casts security functions as SSS services. This strategy allows well-defined protocols and interfaces to be defined for these services and permits an application to outsource security functionality by using a security service with a particular implementation to fit its current need.

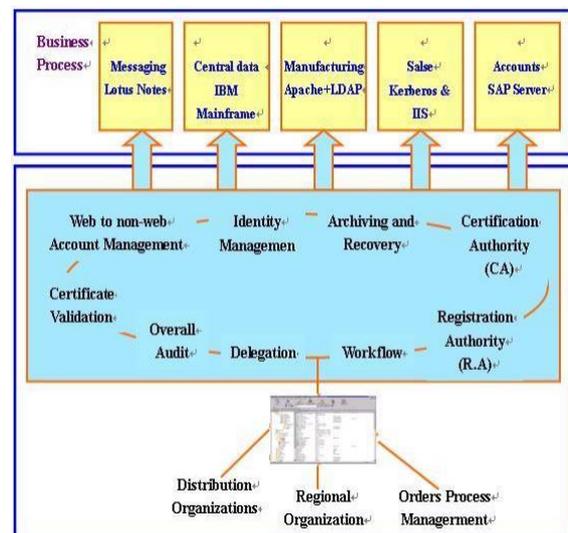


Figure 3 SSS for RFID enabled SCM

The security services include the following.

- ò Certificate Validation service (CAS): A service that handles the details of processing and validating the binding among the subject identity, the subject public key, and subject attributes.
- ò Authorization service: A service that evaluates policy rules regarding the decision to allow the attempted actions based on information about the requestor, and details of the request.
- ò Identity Management service: A service that takes a user's identity in one domain and returns the identity in another (e.g. between X.509 and Kerberos).
- ò Archiving and Recovery: A service that archives and recovers encryption keys.
- ò Administration Service: provide administration workflow with approval and escalation controls.
- ò Overall Audit: A service that securely logs relevant information about events.

For example, SSS perform path validation on a certificate chain according to the local policy and with local PKI facilities, such as certificate revocation (CRLs) or through an OCSP(Online Certificate Status Protocol). The SSS client generates an ‘SSS validate’ request. This is essentially asking the SSS server to go and find out the status of the server’s certificate. Then the SSS server performs a series of validation tasks. After certificate status is determined, SSS server replies to client application with status of the server’s certificate and application acts accordingly. We use CA database connection protocol for the purpose of that the server obtains real-time certificate status information from CAs. The client use OCSP protocol.

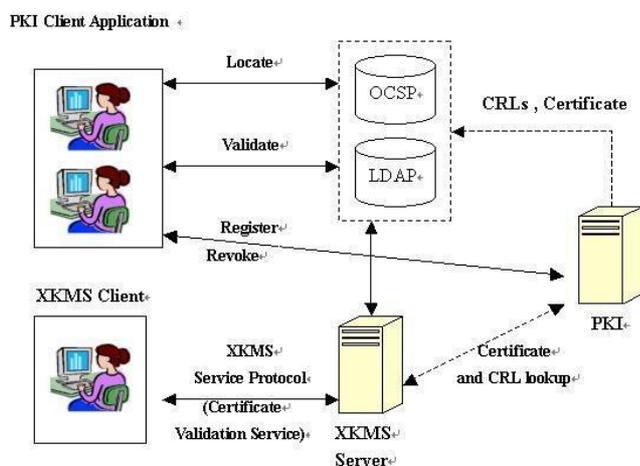


Figure 4 Identify Service

4. PMI IMPLEMENTATION

There are three main components to the PMI implementation – the authorization policy, the privilege allocator (PA) and the PMI application programming interface (API).

4.1 Authorization Policy

The authorization policy specifies who has what type of access to which targets and under what conditions. The xml.org published a document type definition (DTD) for the X.500 PMI RBAC policy. It provided a meta-language that holds the rules for creating the XML policies. The DTD comprise the following components.

- ò SubjectPolicy
- ò RoleHierarchyPolicy
- ò SOAPolicy
- ò RoleAssignmentPolicy
- ò TargetPolicy
- ò ActionPolicy
- ò TargetAccessPolicy

The SOA creates the authorization policy for the domain using his or her preferred XML editing tool and stores this in a local file to be used later by the PA.

4.2 Privilege Allocator

The SOA uses the PA to allocate roles to users in the form of role-assignment ACs. In the simple case of RFID supply chain, there are two roles – ASN issuer and ASN receiver. Issues can download any ASN that are produced.

Once the role-assignment ACs have been created by the PA, they are stored in a LDAP directory.

Another function of the PA is to create an authorization policy that is signed digitally as a policy AC. The policy AC is a standard X.509 AC. The PA prompts the SOA for the name of the policy field and then it copies the contents into the attribute value. After the SOA has signed the policy AC, the PA stores it in the SOA’s entry in the LDAP directory.

4.3 API

A standard authorization API has already been defined by the Open Group in 2000. It is called the AZN API and is specified in the C language. It is based on the ISO 10181-3 access control framework and specifies the interface between the access control enforcement function (AEF) and the access control decision function (ADF).

We rewrite the AZN API with Object-Oriented method, and select Java language for development. After these works, the application gateway use API to “construct” and PMI Java Object, and complete the work by invoke the methods of this object. The main methods include: GetCredits, Decision, and Shutdown.

When a user initiates a call to the target, the AEF authenticates the users, then passes LDAP DN to the ADF through a call to GetCredits. The ADF uses this DN to retrieve all role ACs of the user from the list of LDAP URIs (“pull mode”). The role ACs are validated against the policy e.g. to check that the DN is within a valid subject domain, and to check that the ACs are within the validity time of the policy etc. Once the user has been successfully authenticated he will attempt to perform actions on the target. At each attempt, the AEF passes the subject object, the target name, and the attempted action along with its parameters, to the ADF via a call to Decision. The Decision method will return “Granted” or “Denied” after executing the checking.

Shutdown method is designed to used in dynamically deploy new policy in domain. The AEF can follow the call to Shutdown with a new Constructor call, and this will cause the ADF to read in the latest policy and be ready to make access control decisions again.

5. EXPERIMENT

The first demo system for SSS has been implemented based on the design described in previous section. Some

open source projects are used in the demo system, like Tomcat, Apache SOAP, openCA, and so on. The security API depends core Java security technology, includes Java Authentication and Authorization Service (JAAS), Java Cryptography Extensions (JCE) and Java Secure Socket Extensions (JSSE).

Security service is complicated system. From the running result, there are many works to do in the future. We should consider the performance on the real Internet environment. We need add more robust integration with exist system. And next step, the private keys are generated in the Tags or cards, they can be used as soon as the user is in possession of his tag or card.

6. CONCLUSION

RFID technology is changing the way organizations manage their supply chain networks. With RFID, inventory data becomes more transparent and more accurate, enabling organizations to speed up delivery, respond more quickly to customer demand. But integrating RFID into existing systems has been prolematic, especially, it let suppliers, manufacturers, and customers become a virtual organization across the unsafe internet. We propose a security approach on open web service to validate certificate based on current web security environment, and a privilege management infrastructure used in collaborative business scenario Our approach will be a model for the future security system that offers security of open web service security.

ACKNOWLEDGEMENT

The author would like to thank AutoID Labs China. This work was supported in part by Ubipass company.

REFERENCES

- [1] SAP RFID Home. <http://www.sap.com/solutions/scm/rfid/>.
- [2] Auto-ID Labs. <http://www.autoidlabs.org>.
- [3] ITU-T Recommendation X.509 | ISO/IEC 9594-8, "The Directory: Publib-Key and Attribute Certificate Frameworks", Draft V4 Feb 23,2001.
- [4] S.Garfinkel: PGP:Pretty Good Privacy. O'Reilly & Associates, 1994.
- [5] D. Barret, R. Silverman, SSH: the Secure Shell, O'Reilly & Associates, 2001.
- [6] C. Ellison, SPKI Requirements, IETF RFC 2692, 1999.
- [7] Mary R. Thompson etc., "CA-based Trust Model for Grid Authentication and Identity Delegation", Oct 2002.
- [8] XML-Signature Syntax and Processing (W3C/IETF Recommendation), Feburary-2002
- [9] XML Encryption Syntax and Processing (W3C Recommendation), 2003
- [10] XML Key Management Specification Version 2.0 (W3C Working Draft), April 2003.
- [11]Security Assertion markup language. <http://www.oasis-open.org>
- [12]eXtensible Access Control markup language, <http://www.oasis-open.org>
- [13]Platform for privacy preferences, <http://www.w3.org/P3P/>
- [14]Web Services Security (WS-Security), <http://www-106.ibm.com/developerworks/webservices/library/ws-secure/>
- [15]ACM Workshop on Role Based Access Control, 1996-2001. see <http://portal.acm.org/portal.cfm> for proceedings.
- [16] Lightweight Directory Access Protocol, RFC 2251, Dec. 1997.
- [17]B.Moore, etc. , "Policy Core Information Model", RFC 3060, Feb 2001.

Student Recognition and Awareness of Information Security in Course Learning from Management Information Systems and Computer Science Classes: An Empirical Investigation

Qinyu Liao, Xin Luo, Kirk Arnett

Department of Management and Information Systems, College of Business and Industry,
Mississippi State University, Mississippi State, MS 39762, U.S.A
ql15@msstate.edu¹, xl96@msstate.edu², kpa1@msstate.edu³

ABSTRACT

A survey of college students from both IS and CS programs was undertaken to understand student recognition and awareness of information security issues in course learning. Comparisons will be made for students before and after taking the security class as well as between the two different majors. The results of the study will demonstrate possible differences in perceptions between students of two majors, and identify security issues which have not been identified so that students and institutions are able to improve academic achievements by better tailored security curriculum and training. The findings will also provide feedback to future employers of the students.

Keywords: information security, assurance, awareness, perception, recognition

1. INTRODUCTION

Information security and assurance is becoming a mission-critical component of today's business enterprises, governments, and academic institutes. Notably, after the 9-11 tragedy, more people, such as government leaders, business managers, IT practitioners, and academic researchers, have realized the crucial importance of information security, which, without proper management and supervision, could trigger unforeseeable catastrophe. As a result, a variety of computer and information security training and courses have surfaced at torrent levels in both academia and industry in a bid to improve and revamp employees' and students' awareness, recognition, and assessment of information security. In industry, companies giving serious consideration to information security now yearn to hiring information security/assurance specialists who are able to help them underpin security control, management, supervision, and centralization.

Accordingly, major colleges or universities are endeavoring to open security-related courses, in either Management Information Systems (MIS) programs in Colleges of Business Administration or Computer Science (CS) programs in Colleges of Computer Science/Engineering or both, to spawn these desired information security specialists. And MIS and/or CS major students are enthusiastically interested in taking these courses in an effort to cater to the increasing market demand. The trend is expected to continue to grow in the years ahead as information security and assurance unveils its very importance in today's computing community.

According to literature, a number of researchers have already designed a variety of security frameworks and indicated and measured the importance of establishing security management and control [9, 10, 15, 16]. However, quite a few researches have shed light on assessment of students' information security awareness and recognition from both MIS and CS perspectives in course learning, despite the fact that both academic divisions offer security-related courses at major colleges or universities in the United States. In addition, employers are eager to discover the fitness between the sophisticated work requirements and college graduates' capabilities of handling security issues in the real world. This research, therefore, focuses on observing and identifying undergraduate and graduate MIS and CS major students' awareness, perception, and recognition with regard to information security in course learning.

This research project, funded by a major southern university's Center for Computer Security Research (CCSR) in the United States, intends to answer three questions:

1. What are the security issues recognized by student majors in CS vs. MIS?
2. Are there any differences in their security management skills after taking the security class?
3. Are there any differences in their security technology skills after taking the security class?

2. LITERATURE REVIEW

2.1 Technical Skills and Managerial Skills for IS Security

As more organizations become so dependent on

computer-based and telecommunications-intensive information systems, security has become a vital element for organizations of all types. The unlimited access by a large, knowledgeable community of end users from desktop, dial-in, and network facilities creates a new and extremely vulnerable environment [10]. Each year, billions of dollars worth of information resources are stolen, vandalized, and otherwise misused due to a growing number of industrial spies, computer "hackers," and even unwitting employees [7]. Additionally, Arce and Levy [2] found that the weakest link has shifted from the technology area to the physical area, the human resource area, the policy area, and even individual operating desktop operating systems.

Information and system security has different meanings in the modern business environment. For some, the concept implies reliability (protection against accidental problems) and defensibility (protection against deliberate misuse). For others, IS security is placed into a people context and takes in to account culture, policies, organizational structures, and operating environments. Security technologies tend to address purely technical issues, such as the use of passwords for authentication, or cryptography to guard integrity, and largely ignore the human involvement [4].

Security experts have adopted security measures into four distinct, sequential activities: deterrence, prevention, detection, and recovery [5, 11, 12]. These four classes of sequential actions have a strong theoretical basis in deterrence theory. That is information security actions can deter potential computer abusers from committing acts that implicitly or explicitly violate organizational policy and lower systems risk [16].

In the corporate world, information security is generally seen as being of interest to the IT department, and so many professionals do not give adequate importance to the security concerns of an organization [3]. Straub [16] pointed out that managerial perception of the system security can influence organization security planning and strategies. To ensure the wholeness and soundness of information systems and organization, a security manager will have to take on the role of maintaining integrity of the organizational infrastructure, not just the technical information systems. Therefore, information system security is not a technical problem but a social and organizational problem because the technical systems have to be operated and used by people.

The theory of computer security offers a formal method for security engineering. It has three components: policy, mechanism, and assurance. The theory implies that to achieve a coherent security architecture, security must be considered from the outset, not as an afterthought. Competence in design for security policy enforcement, testing for security, and assessment of security must be part of the education of system

implementers [8]. This agrees with Straub's model for effective system security management.

2.2 Importance of Security Awareness

Security awareness programs are a form of deterrent countermeasure which deserves special mention because educating users as well as their superiors about security yields major benefits [16]. In a survey about MIS management perceptions of the organization security threat, Loch et. al. [10] found out that there is a lack of awareness of the security risks and employee education regarding prevention measures. Managerial awareness is the key to successful security management [16]. To heightened security awareness on the part of managers, professionals, technicians, contractors, etc., all relevant groups should be provided with sufficient training and supporting reference materials to allow them to properly protect and otherwise manage information assets. There are needs to look to organizational theory, management science, and the developing field of information systems instead, for aid and succor in combating threats to security.

2.3 Need for Security Curriculum

Information security jobs in networking and databases will be hot in the next three to five years according to a IT workforce survey in 2004 by Information Technology Association of America [1]. Yet the number of skilled practitioners of computer security who are able to address the complexities of modern technology and are familiar with successful approaches to system security is very small [8]. The recent US Presidential Commission on Critical Infrastructure Protection recommends developing education on methods of "reducing vulnerabilities and responding to attacks on critical infrastructures." The commission recognizes the need to make the "required skills set much broader and deeper in educational level for computer scientists, network engineers, electronic engineers, and business process engineers." By moving to an educational system that cultivates an appropriate knowledge of security, we can increase the likelihood that our next generation of IT workers will have the background needed to design and develop systems that are engineered to be reliable and secure.

When talking about security in engineer education, Irvine [8] gave two important criteria for selecting outcomes for information security education: The education must result in graduates prepared for the security challenges they will encounter in their professional roles; the specific education outcomes for security in a given educational program must be consistent with those of the larger engineering context. Vaughn also considered that the computer science community has an obligation to train its graduates in known protection requirements, causes, vulnerabilities, current research, needed research, and to some extent,

computing ethics so that the graduates could be prepared to defend themselves and their employees from common threats that are known today and to help them understand the problem sufficiently such that they can contribute to discovery of the future solutions that are yet to be found [18].

It has been suggested that security insights must be integrated within the existing information systems programs, rather than be treated separately. There are two possible approaches: computer security could be the focus of the curriculum, which would investigate the foundations and technical approaches to security in considerable depth; a computer science or computer engineering curriculum could choose to use computer security as an important property to be addressed in all coursework [8]. Therefore, it is rare to find computer security course offerings at most academic institutions today. A comprehensive pedagogical approach would suggest that current CS course content be modified to include course specific discussion related to security in networks, database, operating systems, architectures, and software engineering followed by a capstone course learned at the end of the program that is specific to information security issues [18]. Topics appropriate to security-oriented curriculum include security policy models; formal methods applied to system specification, development, and analysis; hardware and software protection mechanisms; security system design, implementation, and testing; database security; modern cryptography; cryptographic protocols; identification and authentication; and coherent network security architecture. Ideally, computer science and computer engineering texts, course materials, and laboratory exercises would have computer security completely integrated into appropriate topics and add a highly relevant dimension to the program- a feature prized by prospective employers [8]. Unfortunately, such materials do not yet exist. The coverage on techniques in security-related textbooks is moderate while other issues like security management have little in depth coverage. Therefore, the classroom coverage of security issue depends on the willingness of instructors to provide supplementary materials [13] because textbooks for courses in computer security provide only a good foundation for the student to learn the techniques that are used with computer systems and networks. In the interim, however, security-related supplements can be used.

Computer science and Information systems are two most common academic subdivisions in the computing field. While computer science examines topics related to computer concepts at technical levels of analysis by formulating processes/methods/algorithms largely using mathematically-based conceptual analysis, information

science topics related largely to organizational concepts, especially usage/operation and technology transfer, although they also explore systems/software topics, all primarily at a behavioral level of analysis.

Glass et al. [6] conducted a study of representative and well-recognized journals from each field over the five-year period 1995-1999, and determined that system security and data security have been given minimal attention in the information systems arena. Reichgelt et al. [14] compared baccalaureate programs in IS, IT and CS. Their study shows that although data security ranked eighth of frequencies for topics offered by the surveyed universities, it is mostly integrated in the curriculum of computer science rather than a separate course. There are very few IS curriculum that cover the security topic.

The Center of Academic Excellence in Information Assurance Education Program was created to help academic institutions prepare security engineers for work in the government IT security area. Universities have responded to this call by setting up separate security course and offering security-oriented course in majors like information system. Students tend to naturally gravitate to this subject area as one of interest and one that offers good skills and knowledge that employers may find attractive [17]. It would be interesting to find out how the security courses are taught in different disciplines by comparing students recognition and awareness of information and system security before and after taking the class. The instructors and institutions can improve the program to fit student background and tailor the course content for future employer needs.

3. RESEARCH METHOD

An empirical investigation will be conducted in *Business Information Systems Security Management* class in the College of Business's MIS program (Course number BIS 6133 for both senior undergraduate and graduate MIS majors) this fall semester starting August 2004. Then data will be collected in the *Information and Computer Security* class in the College of Computer Science and Engineering's CS program (Course number 6243 for both senior undergraduate and graduate CS majors) the following spring semester in 2005 at a major southern university in the United States.

The security technologies and security management skills collected from prior security studies are listed in table 1. These skills are either currently in use or have been proposed by security researchers/practitioners for its importance in computer/system security.

Table 1. List of Issues in Security Technology and Security Management

Security technology	Security management
1. Anti-virus/Malware	1. Disaster recovery
2. Access-control	2. Data backup
3. Firewall	3. Security management policies
4. Password/Log-in	4. Employee education
5. Encryption	5. Patch management
6. Smart card	6. Identity management
7. Authentication	7. Vulnerability management
8. Intrusion detection	8. Procedure of log management
9. Virtual private network (VPN)	9. Procedure of incident report
10. Wireless security	10. Procedure of system maintenance
11. Contents filters	11. Procedure for approval of changes
12. Digital certificate	12. Documentation of changes
13. Network sniffers	13. Budget control
14. Biometric	

Survey questionnaires will be administered prior to and after these two security courses in order to measure any significant change on students' recognition and awareness of information security from both MIS and CS perspectives. Student demographic information such as gender, age, years of computer usage, ownership of computer and experience of other security courses will also be collected for analysis.

4. POSSIBLE FINDINGS

Since we only have data from one out of the four data collections designed for the study, there is no data analysis to report at this point.

A significant limitation of this study is the sample used. Both classes that will be investigated are from the same university and are comparatively small in size because of the availability of separately offered security courses. Also, both are single section courses and are taught on different terms at different times. Future studies should also include students in other universities and courses which has security as an integrated part of the course.

The final result of the study will identify how students' skills and/or perceptions of information security management and technology change after finishing the class. The research findings will be used to help students and instructors achieve further academic excellence in both colleges by fine tuning security curriculums and training and by providing business employers with more instrumental feedbacks with respect to efficient hiring of MIS and CS graduates in the job market.

ACKNOWLEDGEMENT

A number of people have generously given valuable time and effort towards this article. The authors appreciate Dr. Ray Vaughn, Professor of Computer Science and Engineering, for the support of The Mississippi State University Center for Computer Security Research (CCSR), and Dr. Merrill Warkentin

and Dr. J. P. Shim, Professors of Management Information Systems at Mississippi State University, for their research directions and suggestions.

REFERENCES

- [1] The annual IT workforce study by ITAA, 2004.
- [2] Arce, I., and Levy, E., "The weakest link revisited", *IEEE Security & Privacy*, Vol. 1, No. 2, pp. 72-76, 2003.
- [3] Dhillon, G., "Information system security management in the new millennium", *Communications of the ACM*, Vol. 43, No. 7, pp. 125-129, 2000.
- [4] Fink, D., "IS security issues for the 1990s", *Journal of Systems Management*, Vol. 46, No. 2, pp. 46-50, 1995.
- [5] Forcht, L., *Computer Security Management*, Danvers, MA: Boyd & Fraser, 1994.
- [6] Glass, R. L., Ramesh, V., and Vessey, I., "An analysis of research in computing discipline", *Communications of the ACM*, Vol. 47, No. 6, pp. 89-94, 2004.
- [7] Hill, L. B., and Pemberton, J. M., "Information security: an overview and resource guide for information managers", *Records Management Quarterly*, Vol. 29, No. 1, pp. 14-25, 1995.
- [8] Irvine, C. E., Chin, S. K., and Frincke, D., "Integrating security into the curriculum", *Computer*, Vol. 31, No. 12, pp. 25-30, 1998.
- [9] Lee, S. M., Lee, S. -G., and Yoo, S., "An integrative model of computer abuse based on social control and general deterrence theories", *Information & Management*, Vol. 41, No. 6, pp. 707-718, 2004.
- [10] Loch, K. D., Carr, H. H., and Warkentin, M. E., "Threats to information systems: today's reality, yesterday's understanding", *MIS Quarterly*, Vol. 16, No. 2, pp. 173-186, 1992.
- [11] Martin, J., *Security, Accuracy, and Privacy in Computer Systems*, Englewood Cliffs, NJ: Prentice-Hall, 1973.
- [12] Parker, D. B., *Computer Security Management*, Reston, VA: Reston Publishing, 1981.

- [13] Prichard, J. J., and Macdonald, L. E., "Cyber terrorism: a study of the extent of coverage in computer security textbooks", *Journal of Information Technology Education*, Vol. 3, pp. 279-289, 2004.
- [14] Reichgelt, H., et al., "A comparison of baccalaureate Program in information technology with baccalaureate programs in computer science and information systems", *Journal of Information Technology Education*, Vol. 3, 2004.
- [15] Straub, D. W., "Effective IS security: an empirical study", *Information Systems Research*, Vol. 1, No. 2, pp. 255-276, 1990.
- [16] Straub, D. W., and Welke, R. J., "Coping with systems risk: security planning models for management decision making", *MIS Quarterly*, Vol. 22, No. 4, pp. 441-469, 1998.
- [17] Vaughn, R. B., Dampier, D. A., and Warkentin, M., "Building an information security education program", *2004 Information Security Curriculum Development Conference, at Kennesaw State University, Kennesaw, GA*, 2004.
- [18] Vaughn, R. B., and Boggess, J., "Integration of computer security into the software engineering and computer science programs", *The Journal of Systems and Software*, Vol. 49, pp. 149-153, 1999.

A Mediator to Integrate Databases and Legacy Systems: The Mediated Data Integration (MeDInt) Mediator

Chaiyaporn Chirathamjaree

School of Computer and Information Science, Edith Cowan University, Perth, 6050, Australia
c.chirathamjaree@ecu.edu.au

ABSTRACT

To interoperate data sources which differ structurally and semantically, particular problems occur, for example, problems of changing schema in data sources will affect the integrated schema. In this paper, conflicts between heterogeneous systems are investigated and existing approaches to integration are reviewed. We propose a new mediated approach employing the Mediated Data Integration Mediator (MeDInt), and wrapping techniques as the main components for the integration of databases and legacy systems. The MeDInt mediator acts as an intermediate medium transforming queries to sub-queries, integrating result data and resolving conflicts. Wrappers then transform sub-queries to specific local queries so that each local system is able to understand the queries. This framework is currently being developed to make the integration process more widely accessible by using standard tools. A prototype is implemented to demonstrate the model.

Keywords: Conflict Resolution, Heterogeneous Databases, Integration, Legacy Systems, Mediation, Wrappers

1. INTRODUCTION

The information required for decision making by executives in organizations is normally scattered across disparate data sources including databases and legacy systems. To gain a competitive advantage, it is extremely important for executives to be able to obtain one unique view of information in an accurate and timely manner. To do this, it is necessary to interoperate multiple data sources, which differ structurally and semantically. In the process of interoperating any two or more database systems, there are critical problems that need to be solved, for instance, some databases are designed from different models, objects which have the same meaning in different databases might have different names, and objects which have the same meaning in different systems might be measured by different units. Furthermore, there are identity conflicts, representation conflicts, scope conflicts, etc [1; 2; 4; 8; 9]. Although several researchers have studied the conflicts and integration of heterogeneous database systems [1; 9; 11; 13; 14; 17], there is still no common methodology for resolving conflicts and integrating such databases. Particularly, few studies have focused on the integration of databases and legacy systems. In legacy systems, the semantics are hidden and hard to determine. In fact, some legacy systems store data to flat files, which are completely different in schematic design from database management systems (DBMSs).

Another significant issue is that almost all research on database integration presents pre-integration approaches using global schema techniques, which require complete integration. All local views are mapped by one global view. This method is convenient for users but it does not operate in the real-time manner because the global view must be created before query processing. As a result when only one object of a local system is

modified, it affects the global schema requiring huge changes [4]. Furthermore, schema and semantic conflicts must be solved in the process of the global schema creation. The more data sources involved, the more difficult such conflicts are to be solved. This research focuses on the database and legacy integrating solution that avoids using the global schema pre-integration approach.

The Mediated Data Integration (MeDInt) Mediator is introduced in an attempt to overcome the above difficulties. It has been developed by focusing on providing a solution to interoperate heterogeneous data sources by transforming both the queries and the data transparently. Furthermore, this approach does not only solve schema and semantic heterogeneities, but also conflicts from different query languages and data models, namely data model heterogeneity.

2. RELATED WORKS

2.1 Conflicts and Resolution

Information from different sources can not be presented to users if it has not passed the process of conflict resolution. In terms of database integration, conflicts are differences of relevant data between component local database systems. The taxonomy of conflicts in this paper is divided into Schema conflicts and Semantic conflicts.

Schema conflicts are discrepancies in the structures or models of heterogeneous database management systems. Naming conflicts [8], Structural conflicts [4; 8; 9], and Identity conflicts fall into this conflict category. Naming conflicts are the synonyms or homonyms of objects in local systems. Structural conflicts are the different uses of data models to represent the same object. Identity

conflicts occurs when the different attributes, as a key, are used to access the same meaning information.

Semantic conflicts are discrepancies in the meaning of related data among heterogeneous systems such as Naming conflicts, Representation conflicts [3; 4], Scaling conflicts [2], Granularity conflicts, Precision conflicts [1], Missing data, Scope conflicts, and Computational conflicts [2]. Naming conflicts are able to occur in data itself as well as in the structure of data. Representation conflicts or Format heterogeneities are the different uses of formats or data types to represent the same meaning objects. The different units of measurement generate Scaling conflicts.

From a survey of the literature, several methods to resolve conflicts have been found. In the case of Naming conflicts, a catalog [7], tables [4], or meta-data repository [1] can be used for maintaining these correspondences. An Object Exchange model [12] is able to transform semantics into simple structures that are powerful enough to represent complex information by using meaningful tags or labels. Kim [7] suggests three ways to resolve different representations of equivalent data: static lookup tables, arithmetic expressions, and mappings. In addition, a formulae has been suggested by Holowczak & Li [4] for converting values in one system to correspond with units in another system. They also introduce Superclasses to encapsulate each component database to create their relationships. Differences in attribute naming are solved by aliases [1; 4]. By using benefits of functions, Hongjun [5] proposes a data mining approach to discover data value conversion rules. Furthermore, independent views can be constructed to solve Structural conflicts. A view neither depends on any specific names nor on changes when schemas are modified [9].

2.2 Integration Approaches

Numerous integration approaches have been introduced throughout the last twenty years to bring about the interoperability among heterogeneous systems. Missier, Rusinkiewicz, & Jin [10] categorise heterogeneity resolution methodologies into four main broad approaches: Translation, Integrated, Decentralised, and Broker based.

Translation approach needs highly specialised translation for each pair of local database systems. Therefore, the number of translators grows up exponentially especially when local systems increase. The development of these ad hoc programs is expensive in terms of both time and money.

In Tight-coupling approach or fully integrated approach, individual schema from multiple data sources is merged by one or more schemas. If only one schema is prepared, it is called a global schema approach.

Otherwise, it is called a federated database approach. The global schema approach allows access of multiple data sources by providing the conceptual global schema as a logically centralised database [6]. Multiple local schemas are consolidated to create the global schema. Users are able to use one database language to query the global schema without understanding any local schemas. Generally, problems of heterogeneity must be resolved in the process of creating the global schema. A major difficulty is the process of creating global schema which thoroughly understands the differences between the independently-designed heterogeneous local schemas, and homogenises such differences [7]. This approach is more difficult when the number of databases increases. Another approach, the federated database, also allows users to query more than one federated schema without knowledge of local data sources. This approach still requires complete pre-integration. The federated schema must be developed before issuing any queries, so any changes in local schemas would affect the federated schema.

Loose-coupling approach [2] or decentralized approach has been introduced in an attempt to resolve the problems arising from tight-coupling approaches by discarding either pre- or partial-integrated global schema. This approach allows users to query local database systems directly without any global schemas by placing the integration responsibility on users. Multi-database manipulation languages, which are capable of managing semantic conflicts through their specification, are provided as query language tools that are able to communicate with the local databases. Users can see all the local schemas and create their own logical export schema from selected schemas relevant to the information they need [3]. However, it requires users to have semantic understanding and to be able to resolve conflicts in creating their schema, which will be numerous with large numbers of data sources. In Broker-based approach, the crucial part is the conflict detector module using shared ontologies, but the process of doing those ontologies is not completely automated.

The limitations of the above integration approaches have led integration technologies towards a new variety of solutions. Various theories have been applied to solve integration problems such as the object-oriented model, knowledge base [11; 14; 16], ontology [13], and modeling [4].

3. THE MeDInt MEDIATOR

The research has introduced a heterogeneous database integration model incorporating a mediator and wrappers as intermediate layers between the application and data sources. The mediator, MeDInt (Figure 1), serves as an information integrator, between the application and wrappers. Generally, mediators are responsible for retrieving information from data sources, for transforming received data into a common

representation, and for integrating the homogenised data [15]. In this model, the MeDInt Mediator acts as an

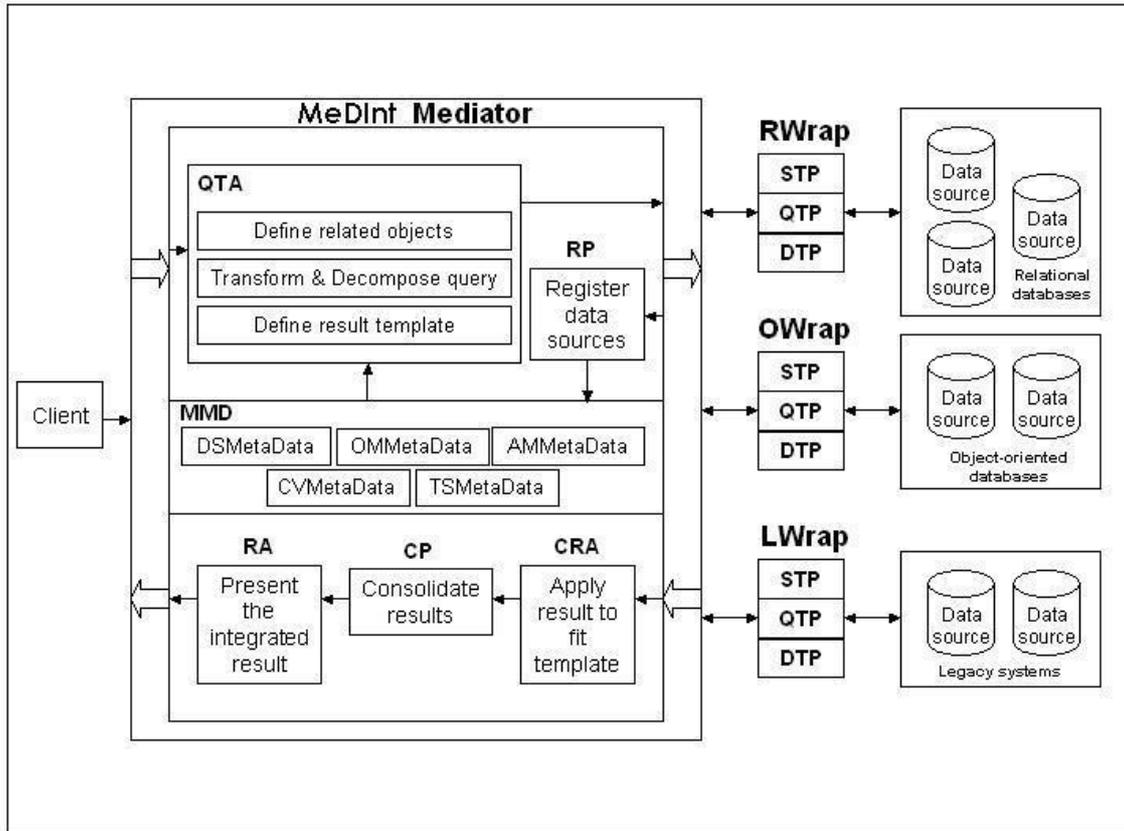


Figure 1. The MeDInt mediator

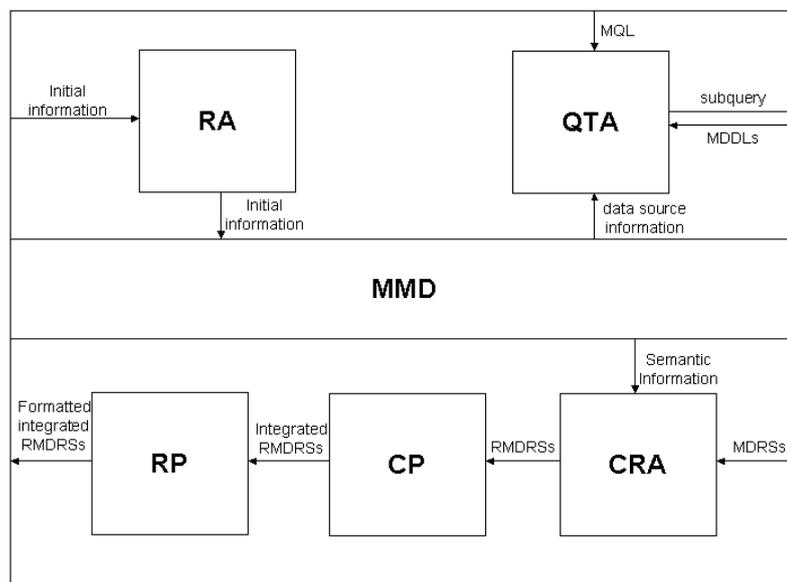


Figure 2. The six components of the MeDInt mediator

interchangeable agent and facilitator for wrappers and clients. It consists of six components working together transparently to facilitate clients and data sources to achieve the following tasks:

- transforming and decomposing the submitted query into subqueries and then distribute them to associated wrappers;
- providing both schematic and semantic knowledge which is critical for query transformation and conflict resolutions;

- resolving conflicts; and
- consolidating query results.

All the functions above are served by six components (Figure 2), which are the Registering Processor (RP), the Query Transformation Agent (QTA), the Mediated MetaData (MMD), the Conflict Resolution Agent (CRA), the Consolidation Processor (CP) and the Rendering Agent (RA).

Wrappers are in the intermediate layer between the MedInt Mediator and data sources. A wrapper is invoked when a data source in a difference data model is added to the integration system. Wrappers mainly act as translators providing the MedInt Mediator with information in the common data model used in the integration system by dealing with the data model heterogeneities of different data sources. The principle objective of wrappers is dealing with data model heterogeneities including the different data definition languages and data manipulation languages by mapping variety data models to the Mediated Data Model. Each MedInt wrapper is composed of a Schema Translation Processor, a Query Translation Processor and a Data Translation Processor. One novel feature of the architecture is to push unshared characteristics the wrappers to reduce the amount of middleware modification when a data source is added, removed or modified. In addition, the use of the Mediated Data Model eliminates problems relating to the data model heterogeneity by providing from the common data model acknowledgeable by components in the MedInt Mediator.

4. Medint PROCESSES

When a new data source is added to the integration system, it is registered to the Mediated MetaData (MMD). Data source information, for example, assigned name, location, type, description, and constraints relating to its structure and semantics are collected into the Data Source Metadata (DSMetaData), a category of MMD. A query from a user to retrieve the information from heterogeneous data sources is sent to the MedInt Mediator instead of directly to the data sources. The required objects are determined and a request is submitted to the wrapper to get the related object schema definitions. The submitted query from the user is transformed to a specific query language appropriate to the database management system of the data source. A template for the results is created from the results obtained from multiple data sources. This method does not try to resolve conflicts directly which would be more difficult and complicated.

After getting a response data back from data sources, a component of a wrapper translates the query results into the Mediated Data Representation Structure (MDRS). The conflict resolution is done by applying all MDRSs to fit into the structure of the predefined template. The

resultant MDRSs that are structurally equivalent are then integrated and consolidated. Finally the integrated result is sent to the user.

This approach overcomes the weakness inherent in other approaches that require the physical or logical integration of component schemas. Only the query result from each source, according to the result template, will be integrated instead. The template will be created from the submitted query. The resultant data from each data source will be applied to fit to the template which is the means by which the heterogeneities are resolved.

5. RESULTS AND DISCUSSION

A number of example problems of heterogeneities from a number of information systems that require integration have been tested. The objectives are to demonstrate the integration process using the MedInt mediator and to evaluate its correctness.

Test problem 1 is a Hotel Reservation Information System which provides information for travel agencies. The information systems of contacted hotels need to be interoperated. Heterogeneities have been found when integrating them. The 2nd test problem is a university information system which is composed of a relational system and an object-oriented system.

The proposed MedInt Architecture and MDM have been tested for functionalities and the outcomes look promising. Results (Table 1) indicate that the objectives in resolving conflicts both structurally and semantically have been achieved. From the table mentioned above, the following three categories of heterogeneities have been determined: Model, Schema, and Semantic. All of them have been solved as shown by the MedInt with the support of the MDM (the Mediated Data Model has been developed in this study specifically for describing and representing heterogeneous data both schematically and semantically) which is suitable for homogenising different data models, schemas and semantics of component data sources. Another feature of our proposed model is that it can be implemented in any languages. We have chosen XML as the implementation language in the prototype because it offers a number of advantages. XML is platform independent, provides self-described tags which are easy to understand. It is also suitable for describing schema and semantic of objects in a real world since XML is based on an object-oriented model.

Table 1 Summary of the heterogeneities resolved by the MeDInt mediator

Heterogeneities	Conflicts	Test Problem1	Test Problem2	
			Query 1	Query 2
Model		✓	✓	✓
Schema	Naming	✓	✓	✓
	Structural	✓	✓	
	Specialisation		✓	✓
	Relationship		✓	
Semantic	Naming		✓	
	Scaling	✓		✓
	Abstraction			✓
	Representation	✓		

6. CONCLUSION AND FUTURE WORKS

The research proposes the MeDInt Mediator as the framework based on the mediated approach for the integration of heterogeneous data sources to solve conflicts occurring when interoperability is required. The paper presents a new approach for achieving the interoperability of multiple data sources logically integrated at the time the query is issued. The system is able to describe or represent heterogeneous data both schematically and semantically. No pre-integration is required before users can issue their queries. This avoids the problem of local schema evolution which usually happens in dynamic systems. Further investigations are planned to cover the query performance issues. Another possible future work is to incorporate the write access through the updating of master data sources and the replication of data sources.

REFERENCES

- [1] Abdulla, K., "A new approach to the integration of heterogeneous databases and information systems", *Unpublished doctoral dissertation*, University of Miami, Florida, 1998.
- [2] Goh, C.H., Madnick, S.E., and Siegal, M.D., "Context interchange: overcoming the challenges of large-scale interoperable database systems in a dynamic environment", *The third International Conference on Information and Knowledge Management*, Gaithersburg, MD., 1994.
- [3] Heimbigner, D., and Mcleod, D., "A federated architecture for information management", in A. Gupta (Ed.), *Integration of information systems: bridging heterogeneous databases*, New York: IEEE Press, 1989.
- [4] Holowczak, R. D., and Li, W. S., *A survey on attribute correspondence and heterogeneity metadata representation*, Institute of Electrical & Electronics Engineers, Available: <http://church.computer.org/conferences/meta96/li/paper.html>, 1996.
- [5] Hongjun, L., *A data mining approach for resolving conflicts during data integration*, Department of Computer Science, The Hong Kong University of Science and Technology, Available: <http://www.comp.polyu.edu.hk/News/Seminars/sem980917.html>, 1998.
- [6] Hughes, J.G., *Object-oriented databases*, New York: Prentice-Hall, 1991.
- [7] Kim, W., *Modern database systems: the object model, interoperability, and beyond*, New York: ACM Press, 1995.
- [8] Kim, W., Choi, I., Gala, I., and Scheevel, M., "On resolving schematic heterogeneity in multidatabase systems", *Journal of Distributed and Parallel Database*, Volume 1, No. 3, p251, 1993.
- [9] Miller, R.J., "Using schematically heterogeneous structures", *SIGMOD'98*, pp189-200, 1998.
- [10] Missier, P., Rusinkiewicz, M., and Jin, W., "Multidatabase languages", in A. Elmagarmid & M. Rusinkiewicz and A. Sheth (Eds.), *Management of heterogeneous and autonomous database systems*, CA: Morgan Kaufmann Publishers, Inc., 1999.
- [11] Neild, T. H., "The virtual data integrator: an object-oriented mediator for heterogeneous database integration", *Unpublished doctoral dissertation*, Northwestern University, 1999.
- [12] Papakonstantinou, Y., Molina, H. G., and Widon, J., "Object exchange across heterogeneous information sources", *ICDE '95 proceedings*, 1995.
- [13] Phijaisanit, W., "Dynamic meta-data support for information integration and sharing across heterogeneous databases (federated database)", *Unpublished doctoral dissertation*, George Mason University, 1997.
- [14] Srinivasan, U., "A framework for conceptual integration of heterogeneous databases", *Unpublished doctoral dissertation*, University of New South Wales, 1997.
- [15] Wiederhold, G., & Genesereth, M., "The conceptual basis for mediation services", *IEEE Expert*, Vol. 12, No. 5, pp38-47, 1997.

-
- [16] Woelk, D., Bohrer, B., Brice, R., Huhns, M., Jacobs, N., Ksieyzk, T., Ong, K., Singh, M., Singh, M., & Tomlinson, C., *Carnot*, Microelectronics and computer technology corporation(MCC), Available: <http://www.mcc.com/projects/infosleuth/archives/carnot/n.d>.
- [17] Yu, T. F., "Information modeling and mediation languages and techniques for information sharing among heterogeneous information systems", *Unpublished doctoral dissertation*, University of Florida, 1997..

A Multi-Attribute Group Decision Approach Based on Rough Set Theory and Application in Supply Chain Partner Selection

Wenqi Jiang, Zhongsheng Hua

Business School, USTC, Hefei, Anhui 230026, China
wqjiang@ustc.edu, wq_jiang@sina.com.cn

ABSTRACT

In multi-attribute group decision, decision makers (DMs) are willing or able to provide only incomplete information because of time pressure, lack of knowledge or data, and their limited expertise related with problem domain, so the alternative sets judged by different decision makers are inconsistent in allusion to a certain decision problem, how to form consistent alternative sets becomes a very important problem. There have been a few studies considering incomplete information in group settings, but few papers consider the adjustment of inconsistent alternative sets. We suggest a method, utilizing individual decision results to form consistent alternative sets based on Rough Set theory. The method can be depicted as follows: (1) decision matrix of every decision maker is transformed to decision table through an new discretization algorithm of condition attributes; (2) we analyze the harmony of decision table of every DM in order to filter some extra alternatives with the result that new alternative sets are formed; (3) if the new alternative sets of different DMs are inconsistent all the same, learning quality of DMs for any inconsistent alternative is a standard of accepting the alternative.

Keywords: group decision; inconsistent; Rough Set; reasoning learning

1. INTRODUCTION

A supply chain is a set of facilities, supplies, customers, products and methods of controlling inventory, purchasing, and distribution. As the global economy has become a reality, only through alliances can firms create more value, it becomes very important for studying the relation of alliances, the management of which is a complicated process in which appropriate partner selection is an important phase. Commonly, decision makers from stock, quality, production, technology and R&D department select partners from many partners with whom they expect to cooperate, some difficult problems exist in partner selection: (1) decision makers from different branches have different preference for the measurement indexes and inaccurate information because of time pressure, lack of knowledge and data, so the partner (alternative) sets offered by different decision-makers will be inconsistent, incomplete and incorrect; (2) different partners and their styles will cause generous number of probable partner combinations, so some alternatives must be thrown off to improve the efficiency of supply chain partner combinations., how to form consistent alternative sets becomes very important. At present, two major approaches are applied in adjusting alternatives: (a) filtering some alternatives using attribute value^[1-3]; (b) configuration learning is used such as artificial neural networks which can not add new rules to the incomplete rule sets^[4-7]. However, the two approaches can't solve the problem of inconsistent alternative sets. Rough set theory^[8], which uses the concept of equivalence classes as its basic principle, was proposed by Zdzislaw Pawlak in 1982 and has been used in reasoning and knowledge acquisition for expert systems.

The idea of this paper can be depicted as follows: (1) the

relative benefit value of all alternatives in a decision matrix are computed through TOPSIS method, an algorithm is presented that condition attributes are discretized with the difference between weight of condition attributes and significance of condition attributes for every patulous decision matrix; secondly, we filter some extra alternatives in decision table by analysis of the harmony of decision table, and the new alternative sets judged by every decision makers are formed, if inconsistent alternative exists in the new alternative sets of different decision makers, learning quality of decision makers for any inconsistent alternative is a standard of accepting the alternative.

2. THE CAUSE OF THE INCONSISTENT ALTERNATIVE SETS

The process of supply chain partner selection includes filtering wildly, filtering carefully, fining and affirming, tracking and appraising. At the stage of filtering wildly, decision-makers from different branches who have different preference for the measurement indexes often make selections from many alternatives offered, and it is difficult for them to gain complete information, so the alternative sets given by different decision-makers are possibly inconsistent^[10-11].

With the hypothesis that the alternative definition is same and decision-makers are $\{DM_i, i = 1, 2, \dots, m\}$ whose alternatives are $A_i = \{A_j, j = 1, 2, \dots, m\}$, for any A_i and A_j , we can describe their relation as follows (α is presented in advance):

$$(1) A_i \setminus A_j = \phi$$

$$(2) \max\left(\frac{\text{card}(A_i | A_j)}{\text{card}(A_i)}, \frac{\text{card}(A_j | A_i)}{\text{card}(A_j)}\right) < \alpha, \text{ more}$$

than $A_i | A_j \neq \emptyset$.

$$(3) \min\left(\frac{\text{card}(A_i | A_j)}{\text{card}(A_i)}, \frac{\text{card}(A_j | A_i)}{\text{card}(A_j)}\right) > \alpha, \text{ more}$$

than $A_i \neq A_j$.

$$(4) A_i = A_j.$$

Our aim is to form consistent alternative sets in favor of group appraisement. This paper applies rough set theory [8-9] which is introduced by Pawlak (1982) to analyze the inconsistent problem in supply chain partner selection and tries to give a method of forming consistent alternative sets, the adjusting process can be depicted as figure 1:

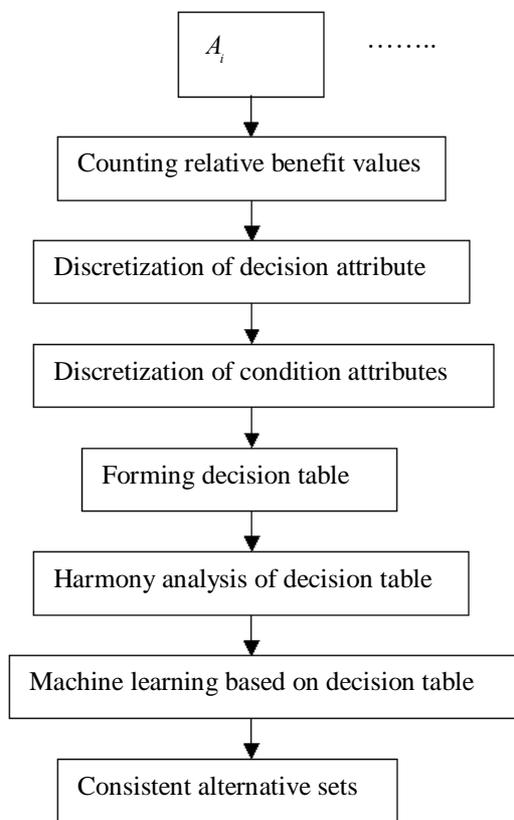


Figure1 : Adjusting procedure figure of inconsistent alternatives

3. ATTRIBUTES DISCRETIZATION

The process of converting data sets with continuous attributes into input data sets with discrete attributes, called discretization, was studied in many papers [12-16], the major methods include the equal-interval-width method, the equal-frequency method, NalveScaler method, SemiNaiveScaler method, etc, but these methods don't contact multi-attribute decision with Rough Set Theory. In the process of multi-attribute decision matrix transformed to decision table, two

difficult problems are put forward that are the unknown value of decision attribute and how to discretize the attributes.

In decision table, the significance of condition attributes reflects weight of condition attributes, by whose difference we can discretize the condition attributes. The procedure can be depicted as follows:

(1) Because the value of decision attribute ($U(a_{ip})$) can be computed by many methods, the paper applies TOPSIS method to compute $U(a_{ip})$ which reflects the near degree between the value of alternative a_{ip} and perfect alternative, the decision matrix and patulous matrix of $DM_i (i = 1, 2, \dots, m)$ can be depicted as table 1 and table 2. We can classify $U(a_{ip})$ into three sorts: $U(a_{ip}) > 0.5$, $U(a_{ip}) < 0.5$, $U(a_{ip}) = 0.5$. The coding are 0, 1, 2 respectively.

(2) The weight of condition attributes of every decision matrix are computed by entropy method, so we can gain

$$w_q(i) = \frac{\ln l_i + \sum_{p=1}^{l_i} b'_{pq}(i) \ln b'_{pq}(i)}{\sum_{q=1}^n \ln l_i + \sum_{z=1}^{l_i} b'_{zq}(i) \ln b'_{zq}(i)}, \text{ the vector of}$$

condition attributes judged by DM_i is $w(i) = (w_1(i), w_2(i), \dots, w_n(i))$.

Table1 : The decision matrix judged by DM_i

	C_1	C_n
a_{i1}	$b_{11}(i)$	$b_{1n}(i)$
a_{i2}	$b_{21}(i)$	$b_{2n}(i)$
.....
a_{il_i}	$b_{l_i1}(i)$	$b_{l_in}(i)$

Table2 : The patulous decision matrix judged by DM_i

	C_1	C_n	$U_i(a_p)$
a_{i1}	$b_{11}(i)$	$b_{1n}(i)$	$U(a_{i1})$
a_{i2}	$b_{21}(i)$	$b_{2n}(i)$	$U(a_{i2})$
.....
a_{il_i}	$b_{l_i1}(i)$	$b_{l_in}(i)$	$U(a_{il_i})$

Where: $\{C_1, C_2, \dots, C_n\}$ is a finite set of condition attributes, $b_{pq}(i)$ shows the value of alternative a_{ip} under the condition attribute $C_q (q \in (1, 2, \dots, n))$. Because the units of different condition attributes may be inconsistent, $b_{pq}(i)$ need be changed into

$$b_{pq}^+(i) (b_{pq}^-(i)) = \frac{b_{pq}(i)}{\sqrt{\sum_{p=1}^l b_{pq}^2(i)}}$$

(3) For condition attribute $C_q, b_{pq}^+(i)$ can be arrayed sort ascending, we give the former class number $m_q = 2$ and apply classification clustering method to classifying the condition attributes into m_q .

(4) We compute the significance of condition attributes of decision table of every decision maker which is $r_c(D) - r_{c-c_q}(D)$, the standardization of the value of $r_c(D) - r_{c-c_q}(D)$ is the effect indexes of condition

attributes which is $v_q(i) = \frac{r_c(D) - r_{c-c_q}(D)}{\sum_{q=1}^n (r_c(D) - r_{c-c_q}(D))}$, if the

significance of all condition attributes is 0, the effect indexes of every condition attribute should be same, which is $\frac{1}{n}$, the vector of the effect indexes can be depicted as $v(i) = (v_1(i), v_2(i), \dots, v_n(i))$.

(5) The difference between the significance of condition attributes and the weight of condition attributes

$$d_{vw}(i) (d_{vw}(i)) = \frac{\sqrt{\sum_{q=1}^n (v_q(i) - w_q(i))^2}}{n}$$

Because the interval spot of condition attributes is finite, the least value exists. In order to reduce to the computing times, we can think of $d_{vw}(i) \leq \beta$ (β is presented in advance) as the last discretization result. If the former discretization can not satisfy the request which is $d_{vw}(i) \leq \beta$, we can find the maximal sensitive degree

$$\max(|\frac{\Delta d_{vw}(i)}{\Delta x_q}|)$$

and change the class of C_q . At last, we also code the condition attributes such as 0, 1, 2, etc, and compute the classifying interval of all condition attributes at the same time.

4. CONFLICTING ALTERNATIVES ANALYSIS IN DECISION TABLE

Decision table which includes many alternatives is a knowledge expression system, a alternative is a decision rule, extra alternatives exist in decision table because of the deficient condition attributes and untrue stylebooks. In table 2, if $\{b_q^+(i)\}$ and $\{b_q^-(i)\}$ denote the perfect

point and imperfect point, $U(a_{ip})$ can be depicted as follows:

$$U(a_{ip}) = \frac{\sqrt{\sum_{q=1}^n (b_{pq}^+(i) - b_q^-(i))^2}}{\sqrt{\sum_{q=1}^n (b_{pq}^+(i) - b_q^+(i))^2} + \sqrt{\sum_{q=1}^n (b_{pq}^+(i) - b_q^-(i))^2}}$$

The value of some condition attributes may belong to the same interval in which if the value of $b_{pq}^+(i)$ tend to

upper end, the value of $\sqrt{\sum_{q=1}^n (b_{pq}^+(i) - b_q^-(i))^2}$ will enlarge and the value of $\sqrt{\sum_{q=1}^n (b_{pq}^+(i) - b_q^+(i))^2}$ will decrease, at last $U(a_{ip})$ will be at different sort.

With the hypothesis of the equation that $A = \bigcup_{i=1}^m A_i$ and $A_i = A \cup A_{i1}$, if some $A_i (A_j | \bigcup_{i \neq j=1}^m A_i) = \emptyset$ exists, a method^[17] is offered to adjust A_j . We can analyze the place of conflicting alternatives, if they belong to A , none alternatives will be deleted; if they belong to A and A_{i1} , we will delete the alternative in A_{i1} ; if they belong to A_{i1} , we will analyze the conflicting alternatives by the significance of condition attributes and delete the alternative whose major attribute value is smaller.

So all decision makers will form new alternative sets $A'_G = \{A'_1, A'_2, \dots, A'_m\}$, where $A'_i = \{a'_{i1}, a'_{i2}, \dots\}$, $A'_i = A \cup A_{i2}$

5. REASONING LEARNING

If the alternative sets A'_i judged by $DM_i (i = 1, 2, \dots, m)$ are inconsistent, the decision makers need reasoning learning which include how to add new learning rules to former decision table and how to compute the learning quality for the new alternative. We can depict the adjusting process as follows:

(1) With the hypothesis of that the inconsistent alternative sets are $\bigcup_{i=1}^m A_{i2}$ which is $A' = \{a_1, a_2, \dots, a_r\}$, and the learning alternative sets chose by DM_i is $\bigcup_{i=1}^m A_{i2} - A_{i2}$.

(2) The weight sets of the decision makers are $\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ we need find the learning decision makers. For example, for inconsistent alternative a_r , the decision makers sets is $\{DM_1, \dots, DM_x\}$, the weight of the DM_t ($t=1, 2, \dots, x$) will change as $\{\lambda'_1, \lambda'_2, \dots, \lambda'_x\}$

where $\lambda'_t = \frac{\lambda_t}{\sum_{t=1}^x \lambda_t}$, the learning attributing value of the

learning decision makers for a_r is

$$\{\sum_{t=1}^x \lambda'_t b'_{t1}, \dots, \sum_{t=1}^x \lambda'_t b'_{tm}\}$$

(3) DM_y ($y=x+1, \dots, m$) collates the learning attributing value as the classifying interval of DM_y and gives the corresponding coding, then the new decision rules come into being.

(4) Computing the learning quality. Considering the relation between learning alternative and intrinsic alternatives, if they conflict, the learning quality can be depicted as follows:

$$k = \frac{card(pos_c(D)) - card(x]_{ind(c)})}{card(U) + 1}$$

where $x]_{ind(c)}$ denotes the equivalence classification including the new example x .

If the example is new, the learning quality is:

$$k = \frac{card(pos_c(D)) + 1}{card(U) + 1}$$

So for alternative a_r , the learning quality

is $f(a_r) = \sum_{y=x+1}^m \lambda'_y f_y(a_r)$, if $f(a_r) > \gamma$ (γ is presented in

advance), the alternatives will be accepted, or the alternative will be deleted.

6. CONCLUSION

This paper analyzes the cause of developing inconsistent alternative sets in multi-attribute group decision, and gives a method of forming consistent alternative sets based on Rough set Theory. The method includes three phrases: attribute discretization, conflicting alternatives deleting in decision table and reasoning learning, and it is applied in supply chain partner selection. In the process of reasoning learning, the paper gives a new rule based on the judgment of the other decision makers, so learning decision makers may get the impact of the other decision makers, how to keep the independence in learning process for the learning decision makers is very difficult, the author

will continue studying the problem.

REFERENCE

- [1] Soung Hie Kim, Chang Hee Han, An interactive procedure for multi-attribute group decision making with incomplete information[J], Computers & Operations Research 26(1999) 755-772
- [2] Soung Hie Kim, Byeong Seok Ahn, Interactive group decision making procedure under incomplete information, European Journal of Operational Research 116 (1999) 498-507
- [3] Bodily, S. E, A delegation process for combining individual utility functions[J], Management Science, 1979, 25(10)
- [4] Towell G G, Shavlik J W, Noordewier M O. Refinement of Approximate Domain Theories by Knowledge Based Neural Networks [J]. Proc of AAAI-90, 1990:861-866
- [5] Towell G G, Shavlik J W, Knowledge Based Artificial Neural networks [J]. Artificial Intelligence, 1994, 70:119-165.
- [6] Optiz D W, Shavlik J W, Heuristically Expanding Knowledge-Based Neural Network [J], Proc of UCAI-93, 1993:1360-1365.
- [7] Fu L Knowledge-based Connectionism for Revising Domain Theories [J]. IEEE Trans on Systems, Man Cybernetics, 1993, 23(1):173-182
- [8] ZDZLSLAW PAWLAK, Rough Sets and Decision Analysis [J], INFOR Vol.38, No 3, Aug 2000
- [9] Salvatore Greco, benedetto matarazzo, Roman Slowinski, Rough sets throy for multicriteria decision analysis, European Journal of Operational Research 129(2001) 1-47.
- [10] Ehap H Sabri, Benita M, Beamon, A multi-objective approach to simultaneous strategic and operational planning in supply chain design, Omega 28 (2000) 581-598
- [11] Christoph Schneeweiss, Distributed decision making in supply chain management, Int J Production Economics 84 (2003) 71-83
- [12] Chiu, D, K,Y, Wong, A K,C, and Cheung, B, Information discovery through hierarchical maximum entropy discretization and synthesis, in Knowledge Discovery in Database(G Piatetsky-Shapiro and W.J.Frawley, Eds), MIT Press, Cambridge, Mass, 125-140, 1991.
- [13] Fayyad, U, M, and irani, K B, On the handing of continuous-valued attributes in decision tree generation, Mach, Learning 8, 87-102, 1992.
- [14] Michal R, Chmielewski and Jerzy W, Grzymala-Busse, Global Discretization of Continuous Attributes as Preprocessing for Machine Learning, International Journal of Approximate Reasoning 1996; 15:319-331
- [15] Bryson, N, Joseph, A, Optimal technidues for class-dependent attribute discretization, Journal of the Operational Research Society, vol52 issue10, p1130
- [16] Susmaga, Robert, Analyzing Discretizations of

- Continuous Attributes Given a Monotonic Discrimination Function, *Intelligent Data Analysis*, Volume1, Issue1-4, 1997, pp157-179
- [17] Satty T L. *The Analytic Hierarchy Process* [M]. New York: Mc Gray-Hill, 1980
- [18] Tzung-Pei Hong, Li-Huei Tsing, Shyue-Liang Wang, Learning rules from imcomplete training examples by rough sets, *Expert Systems with Applications* 22 (2002) 285-293
- [19] A M Malyscheff, T B Trafalis, S Raman, From support vector machine learning to the determination of the minimum enclosing zone, *Computers & Industrial Engineering* 42 (2002) 59-74.
- [20] Ernest Davis, Two machine learning textbooks: An instructor's perspective, *Artificial Intelligence* 131 (2001) 191-198.
- [21] J L Castro, J J Castro-Schez, J M Zurita, Use of a fuzzy machine learning technique in the knowledge acquisition process, *Fuzzy Sets and Systems* 123 (2001) 307-320

A New Clustering Algorithm for Categorical Attributes

Chunbin Tang¹, Weidong Zhao²

¹ Management School, Fudan University, Shanghai 200433, China

² Software School, Fudan University, Shanghai 200433, China
bingotang@hotmail.com, wdzhao@fudan.edu.cn

ABSTRACT

Clustering over categorical attributes is an important yet tough task. In this paper, we present a new algorithm K-means II to extend the famous K-means algorithm which is efficient only on numerical clustering, by using new cluster center definitions and new similarity measures. Thus, our algorithm can be used in categorical clustering while preserving the efficiency. Experiments on both real-life datasets and synthetic datasets show that the K-means II algorithm can produce high quality results and deserve good scalability at the same time.

Keywords: clustering, categorical attributes, similarity, data mining

1. INTRODUCTION

Clustering is a widely used technique in which data points are partitioned into groups, in such a way that points in the same group, or cluster, are similar, and are dissimilar otherwise ^[1]. Much of the previous work focuses on numerical data which can be exploited to naturally defined distance functions between points. There are many algorithms proved to work quite well under numerical conditions, such as K-means ^[2], and so on.

However, much of the data in practice are categorical, where attribute values can not be naturally ordered as numerical values. Clustering for categorical attributes is thus an important task: it is applicable in different domains, e.g. an e-business recommender system. Let's consider a market basket database in a recommender system containing one transaction per customer, each transaction containing the set of items purchased by the specific customer. The transactions can be viewed as tuples with their attributes not numerical but categorical. As is shown in [3], we can set the attribute value as True if and only if an item is purchased; otherwise, it is False. We can cluster over the dataset and get the characterizations of each group, which can be used in targeted marketing and advertising. The characterizations are also effective in predicting buying patterns of new customers based on their profiles. Another example of categorical attributes may be "color" which can get values from the domain {black, yellow, white}. Something should be noted, that not all values of categorical attributes are of interest in practice during clustering. For example, we do not pay much attention to attribute value "False" while clustering on the market basket database.

Categorical clustering is yet a tough work, because the former familiar clustering algorithms for numerical attributes cannot be used directly. Fortunately, several algorithms have been proposed in this domain.

ROCK ^[3] is an adaptation of an agglomerative hierarchical clustering algorithm. They define "links" as the sum of number of common neighbors between two tuples, and then optimize the "links" based criterion function. We can get good results by using this algorithm. However, with the increase of the dataset size, scalability will degrade.

K-modes ^[4, 5] is an algorithm extending the K-means algorithm to work for categorical attributes. They define a new dissimilarity measure for categorical objects, "modes", instead of means, and then use a frequency-based method to update modes in the clustering process. It is scalable as the number of clusters and the dataset size increases. However, the algorithm is unstable because of non-uniqueness of the modes.

STIRR ^[6] is an iterative algorithm based on non-linear dynamical systems. Some research shows that the known dynamical systems cannot guarantee convergence ^[7].

There are a few other clustering algorithms trying to solve the categorical clustering problem from different perspectives, such as CACTUS ^[8], Squeezer ^[9], COOLCAT ^[10], and so on. In this paper, we present K-means II, a new clustering algorithm for categorical attributes. Similar to K-modes, K-means II also extends K-means to categorical space. It modifies the definitions of cluster centers and similarity measures. Like K-means, it can achieve both high quality results and scalability.

The rest of this paper is organized as follows. In Section 2, several definitions related to categorical clustering are given. Section 3 shows our new algorithm in details. Some basic analysis is given in this section as well. In Section 4, experimental results on both real-life datasets and synthetic datasets are demonstrated. Finally, the conclusion is in Section 5.

2. DEFINITIONS

$$\delta_{i,j} = P_j(x_{i,j}) \tag{2}$$

In this section, several definitions related to categorical clustering are given. Our definitions are quite like those in [5].

Definition 1: Let A_1, A_2, \dots, A_n be n attributes describing a space Ω , and D_1, D_2, \dots, D_n the domains of the attributes respectively. D_j is defined as categorical if it is finite and unordered, e.g., for any $a, b \in D_j$, either $a=b$, or $a \neq b$. A_j is then called a categorical attribute. Ω is a categorical space if all A_1, A_2, \dots, A_n are categorical.

Definition 2: A_j is called Boolean attribute if A_j is categorical and has two different values. For simplicity, we often represent D_j as $\{0, 1\}$. Ω is called a Boolean space if all A_1, A_2, \dots, A_n are Boolean. The market basket dataset above is a case in point.

Definition 3: Let the dataset D be a set of tuples where each tuple $t: t \in D_1 \times D_2 \times \dots \times D_n$. If Ω is a categorical space, each data object X in the dataset D can also be represented as a conjunction of attribute-value pairs: $(A_1 = x_1) \cap (A_2 = x_2) \cap \dots \cap (A_n = x_n)$, where $x_i \in D_i$, for $i=1, 2, \dots, n$. For the sake of simplicity, we present X as a tuple: $(x_1, x_2, \dots, x_n) \in D_1 \times D_2 \times \dots \times D_n$.

3. K-MEANS II ALGORITHM

In this section, we introduce a new clustering algorithm over categorical space -- K-means II, which is named after the famous K-means clustering algorithm. Since our K-means II is similar with K-means, two main problems are met in our algorithm as well, which are the calculations of the cluster centers and the similarity measures between tuples and cluster centers. Therefore, we will solve the two problems before putting out our new algorithm.

3.1 Cluster Centers

We now define the center of a cluster as follows. Let $C = \{X_1, X_2, \dots, X_m\}$ be a cluster of data objects, with $X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,n})$, for $i=1, 2, \dots, m$. Let $f_j(a)$ be the number of the tuples with their j -th attributes' value as "a" in the cluster. Define $P_j(a)$ as the percentage of a, thus $P_j(a) = f_j(a)/m$. Furthermore, let P_j be the conjunction of $P_j(a)$, for all $a \in D_j$. Finally, we can define the cluster center Q as follows:

$$Q = (P_1, P_2, \dots, P_n) \tag{1}$$

3.2 Similarity Measurement

Before we define the similarity between the tuple X and the cluster center Q , we first define the similarity between the j -th attribute of X and the j -th attribute of Q as:

The following definition shows the similarity δ_i between X_i and Q :

$$\delta_i = \sum_{j=1}^n \delta_{i,j} / n \tag{3}$$

δ_i works quite well when all values in D_j ($j=1, 2, \dots, n$) are of great importance in clustering the dataset. However, in practice, we sometimes do not pay much attention to some attribute values. As is shown in Section I, the attribute value "False" is not of interest to us in the market basket. Therefore, we should modify the above definitions.

Let V_j be the set of important values of A_j therefore, $V_j \subseteq D_j$. Then we define the similarity $\delta'_{i,j}$ between the j -th attribute of X and the j -th attribute of Q and its weight $\theta_{i,j}$.

$$\delta'_{i,j} = \begin{cases} P_j(x_{i,j}), & x_{i,j} \in V_j \\ 0, & x_{i,j} \notin V_j \end{cases} \tag{4}$$

$$\theta_{i,j} = \begin{cases} 1, & x_{i,j} \in V_j \\ \sum_{a \in V_j} P_j(a), & x_{i,j} \notin V_j \end{cases} \tag{5}$$

Thus, we can get the similarity δ'_i between X_i and Q :

$$\delta'_i = \sum_{j=1}^n \delta'_{i,j} / \sum_{j=1}^n \theta_{i,j} \tag{6}$$

When $V_j = D_j$ (for $j=1, 2, \dots, n$), or in other words, when all values in D_j ($j=1, 2, \dots, n$) are of great importance in clustering the dataset, the two similarity δ'_i and δ_i equals. Thus, we can use δ'_i to replace δ_i . Finally, we get the function $\text{Sim}(X, Q)$ to compute the similarity between tuple X and the cluster center Q as follows:

$$\text{Sim}(X, Q) = \delta'_i \tag{7}$$

The above formula $\text{Sim}(X, Q)$ seems to be quite complex and odd. However, it does derive from the famous simple match coefficient and Jaccard's coefficient* developed to show the similarity for binary data. The function $\text{Sim}(X, Q)$ extends the above two

* The two are coefficients developed for binary data. Let a =conjoint presence (1, 1), b =mismatch (1, 0), c =mismatch (0, 1), and d =conjoint absence (0, 0). Then Simple Matching Coefficient= $(a+d)/(a+b+c+d)$. Absence and presence as well as matches and mismatches have equal weights in this coefficient. Jaccard's Coefficient= $a/(a+b+c)$. Conjoint absence (0, 0) is ignored in the coefficient.

functions, and can be used not only for binary data.

3.3 Our Algorithm

Our K-means II algorithm is similar to the famous K-means [2] algorithm. It can also be described in four steps.

Step 1: Select k cluster centers randomly. (We may choose the first k tuples as the beginning cluster centers)

Step 2: For each tuple X_i , calculate the similarities between X_i and Q_l which is $Sim(X_i, Q_l)$, for $l=1, 2, \dots, k$. Assign X_i to the cluster Q_l (from the formal cluster Q_l) so that the similarity between X_i and Q_l is the largest.

Step 3: Calculate the new k cluster centers, one for each cluster. Update both Q_l and Q .

Step 4: Go to Step 2 until no tuple has changed clusters after a full cycle test of the whole dataset. Otherwise, end.

3.4 Selection of k

The selection of k plays an important role in the final cluster result. To facilitate the selection, we define a notion Δ to show the average similarity of the whole dataset.

$$\Delta = \frac{\sum_{X_i \in D} Sim(X_i, Q)}{num} \tag{8}$$

Q is the center of the cluster which X_i is in, and num is the total number of tuples in the dataset. It is not surprising that Δ increases as k increases. If Δ is too small, it indicates that k is not chosen properly, and we should increase k. So it seems that we will get accurate results if we set k at a high level. However, if k is too large, we will not get very good cluster results, either. We can use Δ as an index to select an acceptable k by experiments.

3.5 Time Complexity

The time complexity of K-means II algorithm depends on the size of dataset (num), the number of attributes (n), the number of iterations (i), and the number of clusters (k). Thus we can get that our algorithm has the worst-case time complexity $O(num * n * i * k)$. It shows that the time complexity is linear with the size of dataset, the number of attributes and the number of clusters. Therefore, we can make the conclusion that our algorithm deserves good scalability.

4. EXPERIMENTAL RESULTS

In this section, the results about the performance of K-means II are demonstrated. We examine the quality of the clustering results on real-life datasets, and the

efficiency on synthetic datasets. Our algorithm is implemented in Java. All the experiments are conducted on a Pentium III-600 machine with 256M of RAM running Windows XP Professional.

4.1 Real-Life Datasets

We experiment our algorithm on two real-life datasets. One is the Congressional Voting dataset, and the other is the Mushroom dataset, both of which can be obtained from the UCI Machine Learning Repository [11]. The two datasets are also used in the algorithm ROCK which can produce good clustering results [3]. Thus, we can compare our K-means II with ROCK.

Congressional Voting Dataset: It is the United States Congressional Voting Records in 1984. Each record corresponds to one Congress man's votes on 16 issues. Nearly all attributes are Boolean (Yes and No) values, and a few contain missing values. A classification label of Republican or Democrat is provided with each data record. The dataset contains tuples for 168 Republicans and 267 Democrats.

Mushroom Dataset: Each tuple in the dataset contains information that describes the physical characteristics of a single mushroom, with a poisonous or edible label. All attributes are categorical. It has 8124 different mushroom records, of which 4208 are edible. A few contain missing values, as is in Congressional Voting Dataset.

Table 1: results for congressional voting dataset

ROCK			K-means II (k=4)		
No.	Rep.	Dem.	No.	Rep.	Dem.
1	144	22	1	23	3
2	5	201	2	140	30
K-means II (k=2)			3	1	34
No.	Rep.	Dem.	4	4	200
1	7	220			
2	161	47			

As Table 1 illustrates, both ROCK and K-means II can identify clusters with the majority of Republicans or Democrats. Due to the elimination of tuples, the sum of the sizes of clusters is not equal to 435 in ROCK. However, K-means II does not eliminate any tuples. As is suggested in sub-section 3.4, we can increase the average similarity Δ by increasing k. Tables 1 also shows the results of K-means II when k = 4. The result is quite accurate and encouraging. Therefore, we can conclude that K-means II can produce good clustering results as ROCK does.

Table 2 describes the clustering results for Mushroom dataset by using ROCK and K-means II. Nearly all clusters found by both algorithms are pure edible or

poisonous ones. Exceptions occur in cluster 5 and 10 (K-means II), and in cluster 15 (ROCK). It shows that both algorithms perform well.

Table 2: results for mushroom dataset

ROCK					
No.	Edible	Poisonous	No.	Edible	Poisonous
1	96	0	12	48	0
2	0	256	13	0	288
3	704	0	14	192	0
4	96	0	15	32	72
5	768	0	16	0	1728
6	0	192	17	288	0
7	1728	0	18	0	8
8	0	32	19	192	0
9	0	1296	20	16	0
10	0	8	21	0	36
11	48	0			
K-means II (k=27, Δ=0.81)					
No.	Edible	Poisonous	No.	Edible	Poisonous
1	192	0	15	224	0
2	96	0	16	0	864
3	288	0	17	0	40
4	372	0	18	192	0
5	16	72	19	0	864
6	276	0	20	324	0
7	0	675	21	32	0
8	192	0	22	324	0
9	0	288	23	432	0
10	96	8	24	128	0
11	768	0	25	0	256
12	0	621	26	160	0
13	0	192	27	0	36
14	96	0			

Experiments on both datasets show that our algorithm can produce accurate clustering results as ROCK does. We can conclude that our algorithm can produce high quality clusters.

4.2 Synthetic Datasets

In order to test the scalability of our algorithm, we experiment with synthetic datasets which are generated by using a data generator. We set the number of attribute values for each attribute to 5. All possible values are produced with (approximately) equal probability. We test three scalabilities of the algorithm using these datasets. The first one is the scalability of our algorithm against the dataset size, the second is the scalability against the number of clusters, and the third is the scalability against the number of attributes. In Figure 1,

the dataset size is increased from 10000 to 100000, and the numbers of attributes and clusters are fixed to 10 and 30 respectively. Figure 2 shows the results when increasing number of clusters from 10 to 50, and the cluster size and the number of attributes are set to 10000 and 10. In Figure 3, we increase the number of attributes from 10 to 40, and we set cluster size at 10000 and the number of clusters at 30 fixedly.

These results are very encouraging because they show clearly a linear increase in time as the cluster size, the number of clusters, and the number of attributes increase. They accord with the analysis in sub-section 3.5. Therefore, we can make the conclusion that our algorithm has good scalability with respect to the above three dimensions.

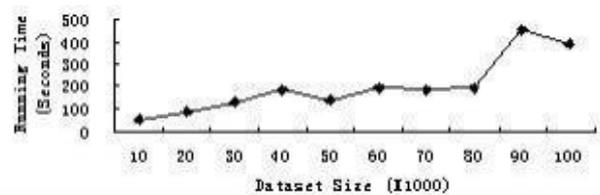


Figure 1: scalability to dataset size (attributes: 10, clusters: 30)

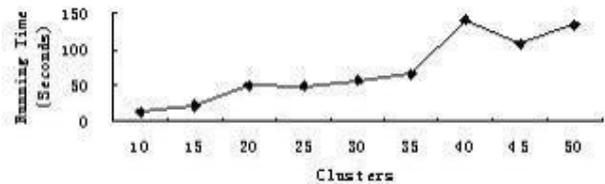


Figure 2: scalability to clusters (dataset size: 10000, attributes: 10)

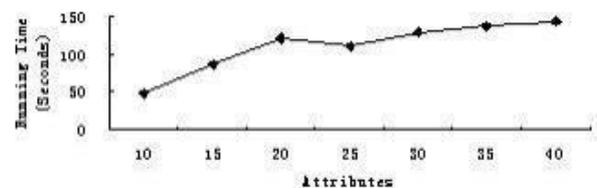


Figure 3: scalability to attributes (dataset size: 10000, clusters: 30)

5. CONCLUSION

Clustering of categorical attributes is an important yet difficult task. In this paper, we propose an efficient algorithm called K-means II. Both analysis and experiments show that our algorithm can produce high quality results and at the same time deserve good scalability.

The K-means II algorithm has made extensions to the K-means algorithm by using new cluster center definitions and new similarity measures. Thus, the K-means II can be used in categorical clustering while preserving its efficiency. The new definition of cluster

center can also be used as characteristic descriptions which are useful in interpreting clustering results. Another advantage is that the algorithm can handle attribute values which are not of interest. Thus, the algorithm can handle missing values very easily, because missing values can be viewed as new attribute values which are not of interest while measuring the similarity.

For the future work plan, we will revise K-means II to make it more efficient on very large dataset size, for examples, datasets with billions of tuples. The extension to K-means II to make it work on mixed numerical and categorical attributes is another direction for us.

REFERENCES

- [1] Anderberg, M., Cluster Analysis for Applications, Academic Press, 1973.
- [2] MacQueen, J., Some Methods for Classification and Analysis of Multivariate Observations, In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, pp281-297, 1967.
- [3] Guhay, S., Rastogi, R., Shim, K., ROCK: A Robust Clustering Algorithm for Categorical Attributes, In Proceeding 1999 International Conference Data Engineering, pp512-521, 1999.
- [4] Huang, Z., Clustering Large Data Sets with Mixed Numeric and Categorical Values, In Proceedings of The First Pacific-Asia Conference on Knowledge Discovery and Data Mining, 1997.
- [5] Huang, Z., A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining, In Proceeding SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, pp146-151, 1997.
- [6] Gibson, D., Kleiberg, J., Raghavan, P., Clustering categorical data : an approach based on dynamic systems, In Proceeding of Very Large Database, 1998.
- [7] Zhang, Y., Fu, A., Cai, C., et al., Clustering Categorical Data, In Proceeding of ICDE, 2000.
- [8] Ganti, V., Gehrke, J., Ramakrishnan, R., CACTUS-clustering categorical data using summaries, In Proceeding of Knowledge Discovery and Data Mining, pp73-83, 1999.
- [9] He, Z., Xu, X., Deng, S., Squeezer: An Efficient Algorithm for Clustering Categorical Data, Journal of Computer Science and Technology, Vol. 17, No. 5, pp611-624, 2000.
- [10] Barbará, D., Couto, J., Li, Y., COOLCAT: An entropy-based algorithm for categorical clustering, Eleventh International Conference on Information and Knowledge Management (CIKM'02), 2002.
- [11] UCI Machine Learning Repository. <http://www.ics.uci.edu/~mllearn/MLRepository.html>

A Personalized Commodities Recommendation Procedure and Algorithm Based on Association Rule Mining

Jianyi Zhang, Yunfeng Wang, Jie Li

School of Management, Hebei University of Technology, Tianjin 300130, China
jyzhang214@eyou.com, {ywang,lijie}@hebut.edu.cn

ABSTRACT

The double-quick growth of EB has caused commodities overload, where our customers are not longer able to efficiently choose the products adapt to them. In order to overcome the situation that both companies and customers are facing, we present a personalized recommendation, although several recommendation systems which may have some disadvantages have been developed. In this paper, we focus on the association rule mining by EFFICIENT algorithm which can simple discovery rapidly the all association rules without any information loss. The EFFICIENT algorithm which comes of the conventional Aprior algorithm integrates the notions of fast algorithm and predigested algorithm to find the interesting association rules in a given transaction data sets. We believe that the procedure should be accepted, and experiment with real-life databases show that the proposed algorithm is efficient one.

Keywords: Association Rules, Data Mining, Procedure, Algorithm, Personalized Recommendation

1. INTRODUCTION

Electronic Business has been growing rapidly, keeping pace with the Web^[1]. According to the investigations of Angus Reid Group in March 2000, there have been more than three hundred million users using the Internet in the world, and there will be three billion ones by 2005^[2]. Its rapid growth has made both companies and customers face a new situation in which companies find it harder to survive due to more and more competition, the opportunity for customers to choose among more and more commodities has increased the burden of information processing before they select which ones meet their needs. In order to solve the problem, researches and practitioners have stressed the need for marketing strategies, for example one-to-one marketing, CRM and mass customization engineering. So under this condition a personalized commodities recommendation which helps our cus-tomers find the ones they prefer are very important and useful.

The great amount of data not only gives the statistics, but also offers the resources of experiences and knowledge. Data mining, as one of the promising technologies since 1990s, is to some extent a non-traditional data-driven method to discover novel, useful, hidden knowledge from massive data sets^[3]. The discovery of interesting association among business decision making processes, such as catalog design, cross marketing, and loss-leader analysis^[4]. Data mining is one of the most popular techniques that can find potential business knowledge from enterprise databases in support of making better decision, it can discovery potentially significant patterns and rules underlying the database. It can be categorized into several interesting areas, such as association rules, clustering, decision tree analysis and so on. Association rule mining finds interesting association or correlation relationships among a large set of data items. *Market basket analysis* is the typical example of association rule

mining^[4].

However, on one hand rules explosion is a problem of concern, as conventional mining algorithms often produce too many rules for decision makers to digest^[5], on the other hand, some rules can miss important information; Addition, the rules of many algorithms are limited to single item. To overcome the shortcomings, we improve on the Apriori Algorithm with EFFICIENT Algorithm. Such a rule set can mine interesting knowledge rapidly without any information loss. At the same time, it can save a large of time and become simpler.

The paper concentrates on the EFFICIENT Algorithm and the recommend procedure of personalize commodities. We begin by show the procedure in section 2. We make clear the basic concept of association rule mining and the EFFICIENT Algorithm in section 3. And section 4 present an example. Finally, we summarize the conclusion and the future research work in section 5.

2. PROCEDURE

Data mining is an interactive and iterative process^[2]. "How are association rules mined from large database?" General speaking, association rule mining includes the following two-step process^[4].

STEP I : Find all frequent itemsets: By definition, each of these itemsets will occur at least as frequently as a pre-determined minimum support count.

STEP II : Generate strong association rules from the frequent itemsets: By definition, these rules must satisfy minimum support and minimum confidence.

In this paper, we only focus on the recommenda tion problem which help customer to find which products they would like to purchase for every one at a specific time; at the same time, the companies can discovery the

rule and recommend related product to their goods. From prepare to discover the relations of the commodities which have been recorded in computers.

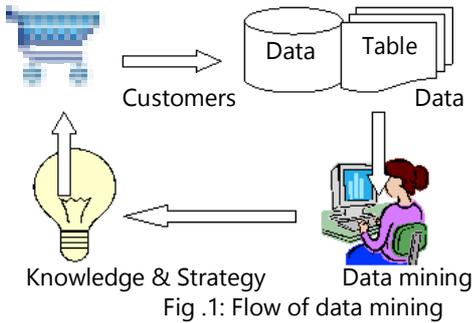


Fig. 1: Flow of data mining

Fig. 2 outlines the procedure of data mining steps. It includes six steps which are cycle [2] [3]

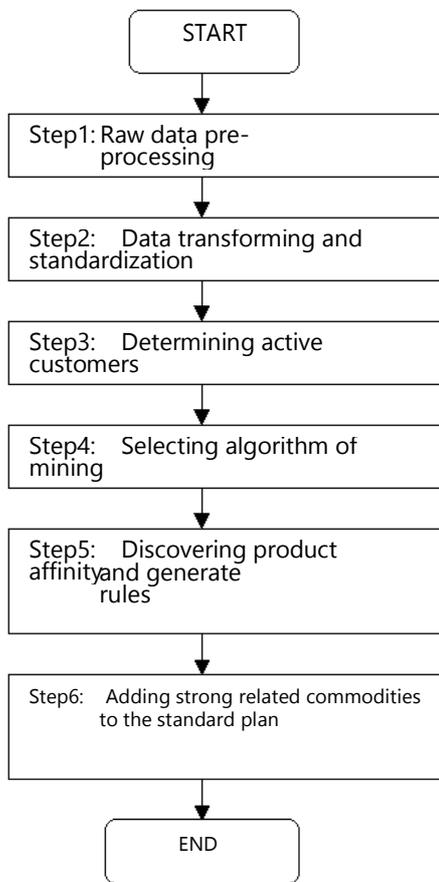


Fig. 2: Procedures of data mining

2.1 Raw data pre-processing

A database consists of current and historical detailed item, summarized data items and metadata. We select database and create a fact table; Next, selecting attributes and dimensions; Finally, we deal with the data, such as filtering data.

2.2 Data Transforming and Standardization

In this step, we transform data and standardize them to

2.3 Determine Active Customers

Making a commendation only for customers who are likely to buy recommended commodities could be avoid the *false positives* of a poor recommendation. In this step, we can use decision tree induction to perform the task of selecting active customers.

2.4 Select Algorithm of Mining

In step4, we should select an appropriate algorithm, although a lot of algorithm can solve mining problem. Some ones have limits which can not sit for many conditions, and some others are complex and invalid. Therefore, we need an efficient algorithm which can fast and simple solve the mining work. EFFICIENT Algorithm which be provided with the character will be referred in section 3.

2.5 Discovering Product Affinity & Generate Rules

In this step, we first search for meaningful relationships or affinities among product classes through mining association rule from large transaction.

Finding association rules at level-k by the following steps and the process of mining will be detailed showed in section4. 1) Get the consequents of rules from l_k with one item in the consequent; 2) Set up minimum support and minimum confidence. At different lever set the *min-supp* and *min-conf* different value. Such is distinctness from the traditional threshold;3) Call *acceleration* (large k- itemset, set of m-item consequents); 4)Call the *predigestion* Algorithm; 5)Obtain the affinity between the stander commodities and target commodities.

2.6 Add Related Commodities

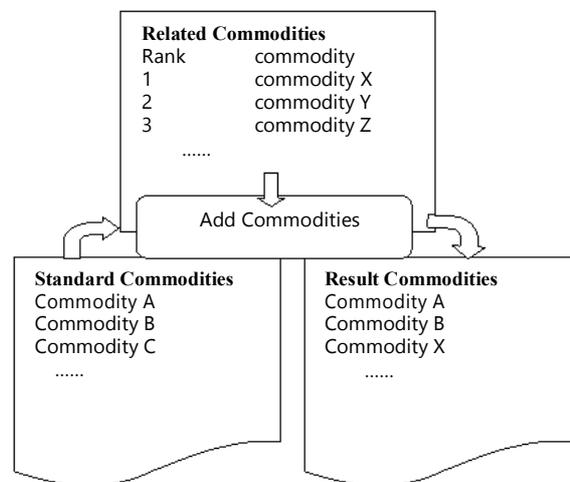


Fig 3: Add commodities to standard plan

We find related commodities are strong related to target commodities, add the related commodities to the standard plan and get the specific sales promotion proposal plan which is prepare to perform the recommendation. This last step can show in fig 3.

The following is the last procedure, but it is the one which realizes the recommend. It is the important phase which verifies the mining wok is useful and efficiency or not.

In fact, the process of mining is a cycle one. When finish a recommendation, the new sale date will be next mined, the consequence still will be doneAt the same time, the rule will be corrected. We can name this process dynamic process.

3. ALGORITHM

"How can we help or guide the mining procedure to discovery interesting association? What language constructs are useful in defining a data mining query language for association rule mining?" [4] In this section, we will find these answers.

3.1 Preliminary Description

Let $J = \{i_1, i_2, \dots, i_n\}$ be a set of items. Let D be a set of database transaction where each transaction T is a set of items such that $T \subseteq J$. Each transaction is associated with an identifier, called TID. Let A be a set of items. A transaction T is said to contain A if and only if $A \subseteq T$. An association rule is an implication of the form $A \Rightarrow B$, where $A \subset J$, $B \subset J$, and $A \cap B = \emptyset$. The rule $A \Rightarrow B$ holds in the transaction in D that contain $A \cup B$. This is taken to be the probability, $P(A \cup B)$ [4].

Rule support and confidence are two measure of rule interestingness. They respectively reflect the usefulness and certainty of discovery rules. The following is the description of them:

$$\text{Support } (A \Rightarrow B) = P(A \cup B) = P(A \cup B) \quad (1)$$

$$\text{Confidence } (A \Rightarrow B) = P(B|A) = P(A \cup B) / P(A) \quad (2)$$

If both the support and confidence are equal, we can use equation (3). If the reverse confidence $(A \Rightarrow B) = 1.0$, when B appear, A must also appear [3].

$$\begin{aligned} \text{Reverse confidence } (A \Rightarrow B) &= P(A|B) \\ &= P(A \cup B) / P(B) \end{aligned} \quad (3)$$

3.2 Apriori Algorithm

Agrawal, Jmielinski, and Swarmi introduced the notion of association rules in 1993. The basic Apriori algorithm regarded as a conventional method, was developed by Agrawal and Srikant (1994) and Uuama Fayyad, and Uthurusamy(1994). Many research efforts have then been made in two directions :one is to extend the notion of

thassociation rules, giving rise to various extensions such as generalized association rules, fuzzy association rules, etc; the other is to improve the algorithm in various ways such as fast algorithms ,sampling algorithm, and so on.

Apriori is an influential algorithm for mining frequent itemsets for Boolean association rules. It employs an iterative approach known as a lever-wise search, where k -itemsets are used to explore $(k+1)$ -itemsets. First, find the set of frequent 1-itemsets. This set is denoted L_1 which is use to find L_2 the set of frequent 2-itemsets used to find L_3, \dots . That all nonempty subsets of a frequent itemset must also be frequent is the Apriori property [4].

The Apriori Algorithm finds the frequent itemsets by using candidate generation. In order to find L_k from L_{k-1} a two-step process which consist join and prune should be done. If $(l_1[1]=l_2[1]) \wedge (l_1[2]=l_2[2]) \wedge \dots \wedge (l_1[k-2]=l_2[k-2]) \wedge (l_1[k-1] < l_2[k-1])$, l_1 and l_2 of L_{k-1} can join, where the condition $l_1[k-1] < l_2[k-1]$ simple ensure that no duplicates are generated. C_k the set of candidates, is a superset of L_k that is, its members may or may not be frequent, but all of the frequent k -itemsets are included in C_k . If any $(k-1)$ -subset of a candidate k -itemset is not in L_{k-1} then the candidate cannot be frequent either and so can be removed from C_k [4][5].

3.3 EFFICIENT Algorithm

EFFICIENT Algorithm is a fast and simple algorithm, which can save a great deal of time without missing useful information. Now let we see many properties of the algorithm [5][9]. First, EFFICIENT Algorithm is simple, in some cases, it is enough, does not need other complex one to solve simple question, such as we can use association rule to class. Second, it count very fast, can decrease the work of CPU, then save significant time. Next, the algorithm pierces into limits of the single item which is required by some algorithms. Finally, generating the focused and interested rules does not need to obtain all rules without miss information.

The flowing is the EFFICIENT algorithm, which includes "acceleration" and "predigestion" actions [3][5][6].

```
// EFFICIENT Algorithm
1) for all large k-itemsets  $l_k, k \geq 2$  do begin
2)    $C_k = \{ \text{frequent } 1\text{-itemsets} \};$ 
3)   If (the value of consequent item = 1)
4)     call acceleration ( $l_k, C_k$ );
5)   else call predigestion ( $l_k, l_k$ );
6) end
//acceleration algorithm
7) Procedure acceleration
   ( $l_k, C_k$  set of  $m$ -item consequents,  $x$ )
8)  $x = x + 1$ ;
9) if ( $k > m + 1$ ) then begin
10)   $C_{m+1} = \text{acceleration}(C_k, x)$ ;
11)  for all  $c_{m+1} \in C_{m+1}$  do begin
12)    conf = sup( $l_k$ ) / sup( $l_k - c_{m+1}$ );
```

```

13)   if (conf < minconf) then
14)     output the rule (l_k c_{m+1}) ⇨ c_{m+1}
        with confidenc=conf and support=supp(l_k);
15)   else delete c_{m+1} from C_{m+1}
16)   end
17)   call acceleration (l_k c_{m+1})
18) end
//predigestion algorithm
19) Procedure predigestion (l_k c_m)
20)   C=[(m-1)-itemsets c_{m-1} | c_{m-1} ⊂ c_m];
21)   for all c_{m-1} ∈ C do begin
22)     conf =supp(l_k)/supp(c_{m-1})
23)     if (conf ≥ minconf)
24)       then begin output the rule c_{m-1} ⇨ (l_k c_{m-1})
        with confidence=conf(l_k) and support=supp(l_k);
25)       if (m-1 > 1) then call predigestion (l_k c_{m-1})
26)     end
27)   end
    
```

Fig 4:The EFFICIENT Algorithm

To generate rules, for every large itemset l, we find all non-empty subset of l. For every such subset c, we output the rule: c ⇒ (l-c), if the support is at least minsupp. The algorithm does not miss any rule because the support of any supplementary item, c' of c must be as great as the support of c. So the confidence of the rule c' ⇒ (l-c') cannot be more than the confidence of c ⇒ (l-c). By rewriting, if rule (l-c) ⇒ c hold, all rules of the form (l-c') ⇒ c' must also hold (c' is non-empty subset of c). If AC ⇒ BD holds, then ABC ⇒ D or ACD ⇒ B must also hold. When call acceleration (ABCDE, ABCE), we can find that ACE ⇒ BD is hold, but ABC ⇒ DE is not [4][8]

Proof: ACE ⇒ BD

$$\begin{aligned}
\because Dconf(X \Rightarrow YZ) &= Dconf(X \Rightarrow Y) * Dconf(XY \Rightarrow Z) \\
Dconf(X \Rightarrow Y) &\geq Dconf(X \Rightarrow YZ) \quad \text{and} \\
Dconf(XY \Rightarrow Z) &\geq Dconf(X \Rightarrow YZ), \quad \text{and} \\
supp(XY \Rightarrow Z) &= Dsupp(X \Rightarrow YZ) \leq Dsupp(X \Rightarrow Y) \\
\therefore XY \Rightarrow Z &\text{ from } X \Rightarrow YZ
\end{aligned}$$

Let X=ACE, Y=B, Z=D, then get ABCE ⇒ D from ACE ⇒ BD, that is ACE ⇒ BD can be expressed by ABCE ⇒ D.

4. EXAMPLE

4.1 Generate Candidate & Frequent Itemsets

This study uses the example to describe the program execution of EFFICIENT Algorithm. The example includes seven transaction records, as show in table 1.

Table 1:The table of database D

TID	List of item	TID	List of item
T1	ABCDE	T5	ABCE
T2	BCE	T6	ACE
T3	ACD	T7	ABCE
T4	BE		

The following is the process of transaction:

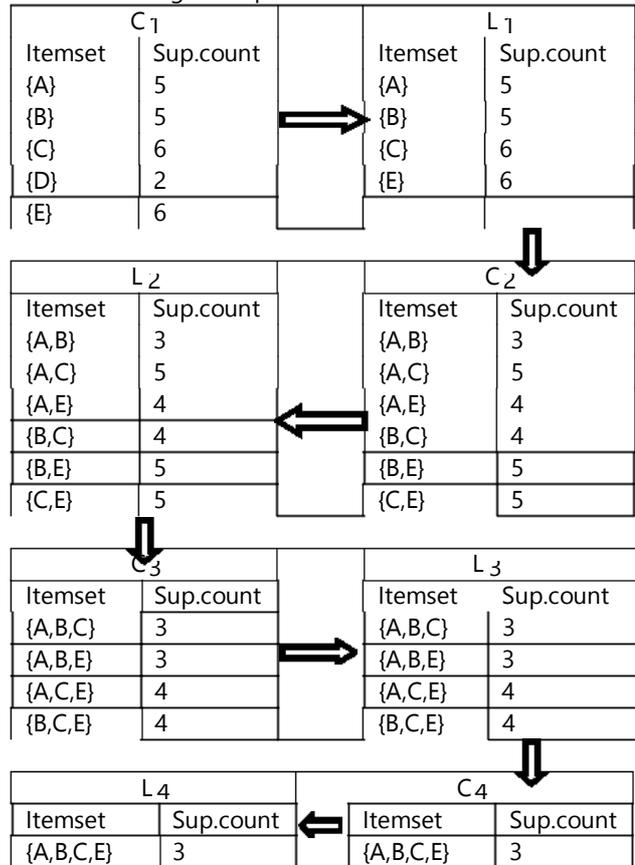


Fig.5: Generate of candidate and frequent itemsets

The minimum length threshold of large itemsets which at different phase is different value, we let the number of items of the first phase be 3 while the min-supp is defined based on the generation of itemsets in various lengths. In the following phases, we can find that the min-supp is changing, that does not appear in additional algorithm. The result is used to prepare the next step which mines the rules.

Table2: All association rules

No.	Rules	Supp	Conf	No.	Rules	Supp	Conf
	limit1	4/7	80%		limit3	3/7	75%
1	A ⇒ C	5/7	100%	14	ABC ⇒ E	3/7	100%
2	A ⇒ E	4/7	80%	15	ABE ⇒ C	3/7	100%
3	B ⇒ C	4/7	80%	16	ACE ⇒ B	3/7	75%
4	B ⇒ E	5/7	100%	17	BCE ⇒ A	3/7	75%
5	C ⇒ E	5/7	83%		limit4	3/7	75%
	limit2	3/7	80%	18	B ⇒ CE	4/7	80%
6	AC ⇒ E	4/7	80%	19	A ⇒ CE	4/7	80%
7	AE ⇒ C	4/7	100%		limit5	2/7	70%
8	CE ⇒ A	4/7	80%	20	AB ⇒ CE	3/7	100%
9	BC ⇒ E	4/7	100%	21	AE ⇒ BC	3/7	75%
10	BE ⇒ C	4/7	80%	22	BC ⇒ AE	3/7	75%
11	CE ⇒ B	4/7	80%		Limit6	2/7	60%
12	AB ⇒ C	3/7	100%	23	A ⇒ BCE	3/7	60%
13	AB ⇒ E	3/7	100%				

Table3: Derived rules

No.	Rules	Derived from
#18	$B \Rightarrow CE$	{3:B \Rightarrow C,9:BC \Rightarrow E}
#19	$A \Rightarrow CE$	{1:A \Rightarrow C,6:AC \Rightarrow E}
#20	$AB \Rightarrow CE$	{12:AB \Rightarrow C,14:ABC \Rightarrow E}
#21	$AE \Rightarrow BC$	{7:AE \Rightarrow C,16:ACE \Rightarrow B}
#22	$BC \Rightarrow AE$	{9:BC \Rightarrow E,17:BCE \Rightarrow A}
#23	$A \Rightarrow BCE$	{19:A \Rightarrow CE,16:ACE \Rightarrow B}
Replaced by		{1:A \Rightarrow C,6:AC \Rightarrow E,16:ACE \Rightarrow B}

4.2 Mine Rule by Using EFFICIENT Algorithm ^{[3][9]}

The following is presented to help illustrate the idea of EFFICIENT Algorithm. If we let the minconf and minsupport have different value, we can get the following table which has some new or different rules.

From the process of mining, we can find that it is more efficient and easier than before. The different between mining for EFFICIENT Algorithm and mining for all association rules is that, to extract the former from any k-frequent itemset (i_1, i_2, \dots, i_k) . The frequent itemset consisting of k-items, one only needs to examine certain frequent itemsets, namely its (k-1)-frequent sub-itemsets instead of all j-frequent sub-itemsets $(j=1, 2, \dots, k-1)$. In fact, only $k(k-1)$ sub-itemsets can be checked, but (2^k-2) sub-itemsets should be done in all association rules.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we introduce a kind of algorithm, EFFICIENT Algorithm which retains all information in the original rule set and save mass time, but has a very smaller size and discovery all significant association rules between items in a large database of transactions. We can draw a conclusion that all qualified association rules can be derived from the EFFICIENT algorithm. The next work is being conducted on further algorithmic optimizations. Besides, more efficient association rules mining may also be the future work. In addition, we will attempt to use the technology of array to solve the mining work, which is a spick-and-span exploration. We believe that data mining is an important new application area for database, combining commercial interest with diverting research questions.

ACKNOWLEDGEMENT

This research was supported by the National Society Science Foundation of China, the Research on the Informationize Model and the System of Information Management oriented MC.

REFERENCES

- [1] Jae Kyeong kim, Yoon Ho Cho, Woo Ju Kim, Je Ran Kim, Ji Hae Suh, "A Personalized recommendation procedure for Internet shopping support", *Electronic Commerce Research and Application* 1(2002)301-313
- [2] S.Wesley Changchien, Tzu-chuen Lu, "Mining association rules procedure to support on-line recommendation by customers and products fragmentation", *Expert Systems with Application* 20(2001)325-335
- [3] Guoqing Chen, Qiang Wei, De Liu, Geert Wets, "Simple association rule (SAR) and the SAR-based rule discovery", *Computers & Industrial Engineering* 43(2002)721-733
- [4] Jiawei Han, Micheline Kambr, DATA MING: Concepts and Techniques, *Higher Education Press*, (2001)225-271
- [5] Rakesh Agrawal, Ramakrishnan Srikant, "Fast Algorithm for Mining Association Rule", *Proceedings of the 20th VLDB Conference*. pp.487-499
- [6] Nicolas Pasquier, Yves Bastide, Rafik Taouil and Lotfi Lakhal, "Efficient Mining of Association Rules Using Closed itemset Lattices", *Information Systems: Vol.24, No.1*, pp25-46, 1999
- [7] Agrawal, R., Imielinski, T., & Swami, A. (1993). "Mining association rules between sets of items in large database". *Proceedings of the ACM-SIGMOD 1993 International conference on Management of Data*, Washington, DC, 207-216.
- [8] Ramakrishnan Srikant, Rakesh Agrawal, "Mining generalized association rules", *Future Generation Computer Systems* 13(1997)161-180.
- [9] A. Savasere, E. Omiecinski, S. Navathe, "An efficient algorithm for mining association rules in large database", 1995, pp432-444.
- [10] Agrawal, R., Mannila, H., Srikant, R., et al. (1996), "Fast discovery of association rules", In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), *Advances in knowledge discovery and data mining* pp.307-328, AAAI Press, Menlo Park

A Practical Approach to Credit Scoring

Jae H. Min¹, Youngchan Lee²

¹ College of Business Administration, Sogang University, Seoul, 121-742, Korea

² College of Commerce and Economics, Dongguk University, Gyeongju, 780-714, Korea

¹ jaemin@ccs.sogang.ac.kr, ² chanlee@mail.dongguk.ac.kr

ABSTRACT

This paper proposes a DEA-based approach to credit scoring. Compared with conventional models such as multiple discriminant analysis, logistic regression analysis, and neural networks for business failure prediction, which require extra a priori information, this new approach solely requires ex-post information to calculate credit scores. For the empirical evidence, this methodology was applied to current financial data of external audited 1061 manufacturing firms comprising the credit portfolio of one of the largest credit guarantee organizations in Korea. Using financial ratios, the methodology could synthesize a firm's overall performance into a single financial credibility score. The empirical results were also validated by supporting analyses (regression analysis and discriminant analysis) and by testing the model's discriminatory power using actual bankruptcy cases of 103 firms. In addition, we propose a practical credit rating method using the predicted DEA scores.

Keywords: Credit scoring, Credit rating, Data envelopment analysis, Financial performance

1. INTRODUCTION

Credit scoring problems are basically in the scope of classification agenda (Anderson, 1984; Chen and Huang, 2003; Dillion and Goldstein, 1984; Hand, 1981; Johnson and Wichern 1998; Lee et al., 2002; Morrison, 1990; West, 2000) that is a commonly encountered decision making task in businesses, and it is a typical classification problem to categorize an object into one of predefined groups or classed based on a number of observed attributes related to that object (Zhang, 2000).

So far, a variety of methods such as linear probability and multivariate conditional probability models, the recursive partitioning algorithm, artificial intelligence approaches, multi-criteria decision-making (MCDM), mathematical programming approaches have been proposed to support the credit decision (Bryant, 1997; Buta, 1994; Coakley and Brown, 2000; Davis et al., 1992; Dimitras et al., 1996, 1999; Emel et al., 2003; Falbo, 1991; Frydman et al., 1985; Martin, 1997; Reichert et al., 1983; Roy, 1991; Tam and Kiang, 1992; Troutt et al., 1996; Zopounidis and Doumpos, 1998).

Offering financial institutions a means for evaluating the risk of their credit portfolio in a timely manner, such models can provide an important body of information to help them formulate their respective risk management strategies. In fact, banking authorities such as Bank of International Settlements (BIS), the World Bank, the IMF, and the Federal Reserve all encourage commercial banks to develop internal models to better quantify financial risks.¹

¹ The Basel Committee on Banking Supervision (1999), English and Nelson (1998), the Federal Reserve System Task Force on Internal Credit Risk Models (1998), Lopez and Saidenberg (2000), and Treacy and Carey (2000) represent some recent documents addressing these issues.

The purpose of this paper is to suggest a new approach to credit scoring, which is based on DEA. As opposed to well-known methods such as multiple discriminant analysis, logistic regression analysis, and neural networks, which require ex ante information of "good/bad" classification, this approach only needs ex post information of the observed set of input and output data of the objects of interest (client firms) to calculate their respective credit scores. With these scores, we also provide a practical credit rating methods to classify client firms into several balanced classes.

2. LITERATURE REVIEW

In the credit industry, neural networks (NN) has recently been claimed to be an accurate tool for credit analysis among others (Desai et al., 1996; Malhotra and Malhotra, 2002; West, 2000). Desai et al. (1996) have explored the abilities of NN and the traditional statistical techniques such as linear discriminant analysis (LDA) and logistic regression analysis (LRA) in constructing credit scoring models. They claimed that NN shows a promise if the performance measure is the percentage of bad loans accurately classified. However, if the performance measure is the percentage of good and bad loans accurately classified, LRA is as good as NN. The percentage of bad loans correctly classified is an important performance measure for credit scoring models since the cost of granting a loan to a defaulter is much larger than that of rejecting a good applicant (Desai et al., 1996).

West (2000) has also investigated the accuracy of quantitative models commonly used by the credit industry. The results indicated that NN could improve the credit scoring accuracy. He also suggested that LRA is a good alternative to NN while LDA, k-nearest neighbor (k-NN), and CART (classification and

regression tree) did not produce encouraging results. Commonly considered as a black-box technique without logic or rule-based explanations for the input-output approximation, the main shortcoming of applying NN to credit scoring lies in the difficulty of explaining the underlying principle for the decision to rejected applications (West, 2000).

Although NN and other traditional methods for credit scoring require ex ante information for business failure prediction, it is more useful in practice to build a credit scoring model based on ex post financial information. The idea is to develop a meaningful “peer group analysis” with specific financial characteristics that distinguish between two or more groups, and in the late 1990s, data envelopment analysis (DEA) was introduced to this peer group analysis for business failure prediction (Troutt et al., 1996; Simak, 1999; Cielen and Vanhoof, 1999).

As opposed to broadly known MDA, LRA, NN approach, DEA requires solely ex-post information, i.e. the observed set of input and output data, to calculate the credit scores. Yeh (1996) was one of the pioneers to combine DEA with financial ratio analysis. She utilized DEA to evaluate bank performance. Her study empirically demonstrated that DEA, in conjunction with financial ratio analysis, can effectively aggregate and reclassify perplexing ratios into meaningful financial dimensions, which enable analysts to gain an insight into the operating strategies of banks. Emel et al. (2003) proposed a credit scoring methodology based on DEA. Although their approach, which is applied to the limited number of Turkey’s commercial banks, is not relatively delicate compared with conventional statistical analyses, it provides the base of this study.

DEA converts a multiplicity of input and output measures into a unit-free single performance index formed as a ratio of aggregated output to aggregated input. Conceptually, DEA compares the DMUs’ observed outputs and inputs in order to identify the relative “best practices” for a chosen observation set. Based on these best observations, an efficient frontier is established, and the degrees of efficiency of other units with respect to the efficient frontier are measured. Therefore, in the context of credit scoring, the performance index via DEA measures the relative credit riskiness of the firms within credit portfolio (Emel et al., 2003). DEA, which computes a firm’s efficiency by transforming inputs into outputs relative to its peers, may provide a fine mechanism for deriving appropriate categories for this purpose.

3. RESEARCH METHODOLOGY

The research methodology consists of seven steps, as outlined in Fig. 1. The first three steps deal with selection of firms for the study and with identification of indicators that may be used to evaluate the firms’

financial performance. Step 4 uses DEA to obtain credibility scores of the firms. Step 5 validates the DEA-based credibility scores by comparing them against those obtained via regression and discriminant analyses, and by using actual bankruptcy cases. Finally, Step 6 proposes a credit rating method by investigating the distribution of good/bad firms’ credibility scores.

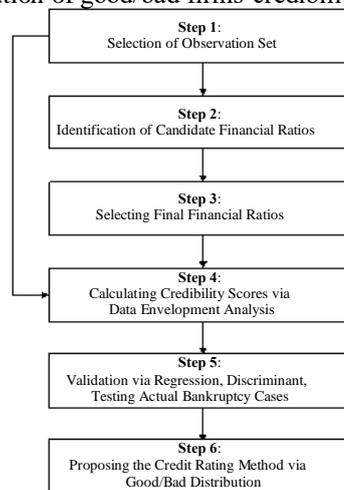


Fig. 1. Flowchart of the research methodology

Step 1: Selection of observation set. We select the firms applying for new credit allocation or whose credit limits are already allocated by the credit authority. At this stage, a certain degree of “homogeneity” in terms of industrial difference and scale-size is ensured among the firms. In this paper, we select only external audited manufacturing firms as a sample in order to satisfy this property.

Step 2: Identification of candidate financial ratios. The most common dimensions considered in financial performance evaluation are growth, liquidity, activity, profitability, productivity, and cost structure aspects. In order to cover these dimensions, a broad set of financial ratios needs to be computed. Some ratios in this set, however, may be similar to each other in terms of underlying financial meanings or in terms of mathematical properties. To identify diverse and financially meaningful ratios in this model, the literature review as well as loan officers’ experience-based insight was employed.

Step 3: Selecting final financial ratios. The final selection of financial indicators is based on the expert opinion as well as the statistical factor analysis. The resulting set of indicators contains the most relevant financial classification dimensions while recognizing the mathematical relationships among the ratios.

Step 4: Calculating credibility scores via data envelopment analysis. In DEA, physical or monetary magnitudes are typically used as the input/output set. However, to eliminate scale-size effects in this study, financial ratios were used instead. The resulting DEA score is a relative ratio of two combined linear ratios. Also, we took advantage of multi-criteria ranking

feature of DEA, a feature based on selection of the relatively best practices within the observation set, and on the radial distance from the efficient frontier comprising these best practices (Charnes et al., 1978; Oral and Yolalan, 1990).

Step 5: Validation via regression, discriminant, and testing actual bankruptcy cases. The purpose of this step is to establish the extent to which DEA results coincide with those of regression analysis, discriminant analysis, and actual bankruptcy cases.

(a) *Regression analysis (RA):* In some cases, due to data anomalies, DEA may not sufficiently discriminate firms' efficiencies. Thus, there is a need to test the explanatory power of the indicator set used in DEA. Linear regression analysis is suggested as a test criterion. For this purpose, the DEA scores are taken as the dependent variable, while the financial ratios used in DEA are set to be the independent variables.

(b) *Discriminant analysis (DA):* DA is used to establish the extent to which DEA scores can be used to classify the sample firms into two groups: "good" firms and "bad" firms. DA is a statistical technique used to classify an observation into one of a priori established groupings dependent upon the observation's individual characteristics. DA attempts to derive the linear combination of characteristics which best discriminates between predefined groups. In this study, financial performance, as measured by DEA, is used as the qualitative (i.e. a priori grouping) variable. There are two performance groups: The good firms group and the bad firms group. The good firms group is defined as those observations with DEA scores over a specific value, whereas the bad firms group is defined as those observations with DEA scores below that value. As will be discussed later, the specific value was chosen by taking into account the distribution of DEA scores. The financial ratios are used as explanatory variables in DA. One can then generate a discriminant function, and calculate the hit ratio (the percentage of right classifications) that shows the degree to which DA validates the classification obtained via DEA.

(c) *Testing actual bankruptcy cases:* The consistency of the DEA results is also checked against testing actual bankruptcy firms. The objective is to see the usefulness of DEA scores as a risk management tool in practical viewpoint.

Step 6: Proposing the credit rating method via good/bad distribution. Commercial banks or other financial institutions are adopting various credit rating methods to manage the client firms' credit riskiness, most of which utilize the probability of default derived by the neural networks or logistic regression analysis. In practice, an ex post approach is more useful in order to diagnose the financial performance of client firms and to rate their credit status. In this study, we propose a practical credit

rating method by investigating the distribution of the firms' credibility scores.

4. EMPIRICAL ANALYSIS

The DEA-based approach was applied to current financial data of external audited 1061 manufacturing firms comprising the credit portfolio of one of the largest credit guarantee organizations in Korea.

4.1 The sample data

At the beginning of the study, there were approximately 1400 firms for which data were available. In order to provide a certain degree of homogeneity among firms in the observation set, however, the outliers, i.e. those firms having several ratios that deviate significantly (more than two standard deviations) from the corresponding mean, were removed and the data for the remaining 1061 firms were used.

4.2 Selection of financial ratios

Commonly accepted financial dimensions such as growth, liquidity, activity, profitability, productivity, and cost structure aspects are considered as a guideline to identify candidate financial ratios. From the previous studies (Altman, 1968; Beaver, 1966; Dimitras et al., 1996; Eisenbeis, 1978; Emel et al. 2003; Falbo, 1991; Jensen, 1992; Lee et al. 1997; Lee et al., 1999; Martin, 1997; Peel et al., 1986), we first selected 57 financial ratios, of which 43 ratios were grouped under previously mentioned dimensions through factor analysis. The loan officers' experience-based knowledge was then used to select final financial ratios that represent a firm's multidimensional financial characteristics. Combining the credit department officers' expert knowledge, the literature survey, and the authors' best judgment, we selected the final set of 6 financial ratios, and classified them as input and output variables for DEA.

The inputs to be minimized are financial expenses to sales, current liabilities ratio and absolute value of 1-fixed assets ratio, as defined in Table 1.

Table 1. Input variables for DEA

Input variables	Formula
financial expenses to sales (FE)	financial expenses ÷ sales
current liabilities ratio (CL)	current liabilities ÷ owners' equity
1-fixed assets ratio (ABS)	1-(fixed assets÷owners' equity)

First, the ratio of financial expenses to sales (FE) shows the ability of a firm to pay financial expenses, which indicates the credit worthiness of a firm. Second, current liabilities ratio (CL) is a proportion of current liabilities to owners' equity. This ratio is an indicator to measure stability of a capital structure. If this ratio becomes

higher, a capital structure and a financial liquidity are in unstable status. Third, the absolute value of “1 minus fixed assets ratio”(ABS) shows that fixed assets of a client firm should balance its capital base. If banks finance fixed assets with liabilities, especially current liabilities (since fixed assets will not bring revenues to the bank, at least in the short run), the client will have problems in paying back the credit. This will also lead to cash flow problems for the bank. Hence, the ratio of fixed assets to capital base should be close to one. As this ratio moves away from unity (1) in either direction, it indicates an imbalance.

The outputs to be maximized are capital adequacy ratio, current ratio and interest coverage ratio, as defined in Table 2.

Table 2. Output variables for DEA

Output variables	Formula
capital adequacy ratio (CA)	owners' equity ÷ total assets
current ratio (CR)	current assets ÷ current liabilities
interest coverage ratio (IC)	(EBIT + interest expenses) ÷ interest expenses

First, the capital adequacy ratio (CA) is a proportion of owners' equity to total assets. This ratio is an indicator of the capital adequacy of the firm. The more a firm finances itself with its own resources (the higher this ratio is), the less risky it is evaluated by credit authorities. Second, the current ratio (CR) is an indicator of the client's liquidity. The more liquid the firm is, the easier it can pay its current obligations. Therefore, the higher this ratio is, the better liquidity position the firm is in. Third, the interest coverage ratio (IC) is an indicator that shows the ability of a firm to pay its interest expenses with operating income. Therefore, the higher this ratio is, the more profitable the firm is.

4.3 Calculating financial credibility scores via DEA

Setting FE, CL, and ABS as input variables while CA, CR, and IC as output variables, we ran DEA algorithm and computed the financial performance (credibility) scores of the 1061 firms. The scores were calculated using input-oriented CCR model assuming constant returns to scale. In this application, DEA scores were given as percentage points. Hence, the range of scores in the original model, i.e. 0-1, will be reported as 0-100.

The resulting DEA credibility scores vary between 13.04 and 100. Firms with DEA score of 100 are considered best firms and are said to fall on “efficient frontier.”Fig. 2 shows the distribution of DEA scores for the 1061 sampled firms.

As seen in Fig. 2, there are 16 firms with DEA scores of 100. As the DEA score of a firm is lower than others, its financial performance is considered relatively worse than other firms in the observation set. It is thus considered to be closer to a probable bad risk in the context of loan extension process.

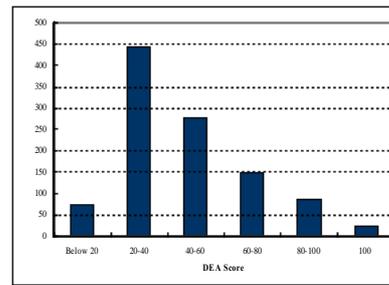


Fig. 2. Distribution of the DEA scores for the 1061 sampled firms

4.4 Validating DEA scores

(a) *Regression analysis:* In this context, DEA credibility score acts as the dependent variable while the six ratios are considered as the independent variables. To prevent overestimation, the regression was run excluding the “best observations”(firms with DEA score of 100).

Table 3. Regression analysis results

R ² = 0.741; F=491.803 (Sig. = .000)					
	Unstandardized Coefficients	Standard Error	t-value	p-value	VIF
Constant	72.94035	3.772205	19.33626	0.000	
FE	-109.653	21.6798	-5.05785	0.000	1.446
CL	-50.8614	3.123891	-16.2814	0.000	1.726
ABS	-56.3548	3.039185	-18.5427	0.000	1.411
CA	47.45822	4.30213	11.03133	0.000	1.790
CR	17.36333	1.085729	15.99233	0.000	1.273
IC	1.146138	0.184531	6.211091	0.000	1.324

As shown in Table 3, all the variables have expected directions and are statistically significant, which tells us that the DEA algorithm successfully accounted for all six ratios at a statistically significant level.

Equation (1) represents the estimated regression relationship. This can be seen as a linear approximation of the DEA results. If the observation set is statistically large enough, the regression equation may also be used to evaluate a new credit applicant without having to run all the steps to derive its DEA score. In other words, by using Equation (1), it is possible to compute the linear approximation of its DEA score without having to run the DEA algorithm each time a new observation is added.

$DEA = 73 - 109.7FE - 50.9CL - 56.4ABS + 47.5CA + 17.4CR + 1.2IC$
 Using the regression equation, we computed the “fitted DEA scores”and compared them with those obtained by DEA. As shown in Fig. 3, the actual DEA scores and the fitted scores do not differ significantly. The matched-pairs t test also assured that there is no significant difference between the actual scores and the fitted ones (t-value=0.000, p-value=1.000).

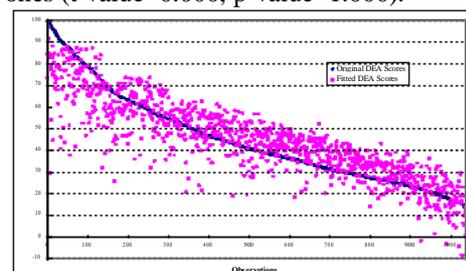


Fig. 3. Actual vs. fitted DEA scores

(b) *Discriminant analysis*: An attempt was made to approximate the DEA results through DA. The firms were classified into two groups with respect to their DEA scores. The “cut-off” point between good and bad firms was selected in an ex post subjective manner, giving due consideration to the distribution of the DEA scores.

Table 4. Discriminant analysis result

		Selected Group		
		Good	Bad	Total
Predicted Group	Good	440 (84.9%)	57 (11.0%)	497
	Bad	78 (15.1%)	461 (89.0%)	539
	Total	518	518	1036

In this study, median of DEA scores is selected as the cut-off point due to their skewed distribution. Thus, 518 firms were classified as good while the remaining ones were classified as bad. Next, DA was run with the above classification scheme as the category variable and the six ratios as the independent variables. The DA generated a discriminant function with five of the six ratios included (only IC being excluded). Table 4 showed that DA resulted in (440+461)/(518+518) or 87.0% hit ratio.

Equation (2) represents the unstandardized canonical discriminant function:

$$Z = -3.2 + 9.0FE + 4.4CL + 5.6ABS - 3.4CA - 1.5CR(2)$$

As shown in Fig. 4, the DA-generated ranking did not differ significantly from that obtained by DEA, which is also statistically ensured by Spearman test (rank order correlation coefficient=0.882, p-value=0.0001). This tells us that the financial credibility scores derived by DEA can indeed be linearly approximated by DA, and provide useful information for business failure prediction.

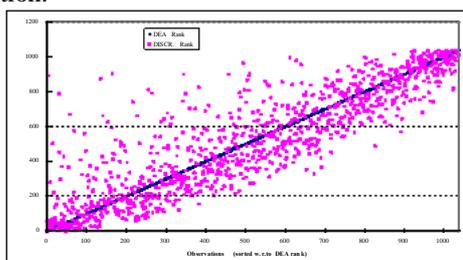


Fig. 4. DEA vs. DA rankings of sampled firms.

(c) *Testing actual bankruptcy cases*: We also checked consistency of the DEA results using actual bankruptcy cases of 103 firms. As shown in Table 5, the hit ratio of bankruptcy classification turned out to be 78.6%. This result shows the usefulness of DEA-based methodology in financial distress prediction while DEA only considers ex post information of the firms.

Table 5. Classification result of actual bankruptcy firms

		Predicted Group		
		Good	Bad	Total
Actual Group	Bad	22 (21.4%)	81 (78.6%)	103

4.5 Credit rating method

From a practical risk management point of view, the fitted DEA scores (via regression analysis) of credit applicant firms should be classified into several classes such as A, B, C, and so on. Table 6 shows the frequency distribution of good/bad firms’ actual DEA scores with equal class intervals.

Table 6. Original frequency distribution of good/bad firms

Class	Good	Bad	Sum	% of Bad	% of Sum
Below 20	76	25	101	24.8%	8.7%
20-40	360	46	406	11.3%	34.9%
40-60	402	27	429	6.3%	36.9%
60-80	173	5	178	2.8%	15.3%
80-100	48	0	48	0.0%	4.1%
Upper 100	2	0	2	0.0%	0.2%
Total	1061	103	1164	8.8%	

As one can see in Table 6, the relative frequency of bad firms (% of Bad) decreases as the DEA scores increase. This means that the DEA score can serve as a very useful indicator to quantify the credit worthiness of the applicant firms, and thus banks or other financial institutions may grade the client firms’ credit according to the DEA score classes. For example, a firm with DEA score ranged from 20 to 40 may be graded “E.” However, the frequency of the firms in each DEA score class (% of Sum) in Table 6 has a drawback in real world applications.

Table 7. Modified frequency distribution of good/bad firms

Interval	Good	Bad	Sum	% of Bad	% of Sum
Below 25	138	37	175	21.1%	15.0%
25-35	186	28	214	13.1%	18.4%
35-45	219	16	235	6.8%	20.2%
45-55	222	15	237	6.3%	20.4%
55-70	177	7	184	3.8%	15.8%
Upper 70	119	0	119	0.0%	10.2%
Total	1061	103	1164	8.8%	

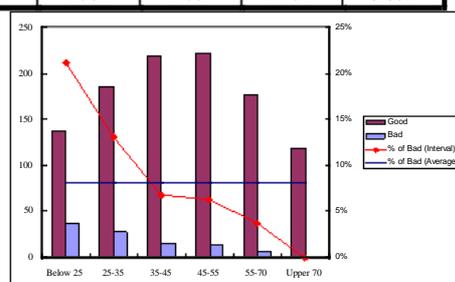


Fig. 6. Histogram of modified frequency distribution

Commercial banks or other financial institutions do not normally grade the client firms as shown in Table 6 due to their respective internal business policies. Table 6 shows the DEA score distribution (% of Sum) is skewed to the right, which indicates that there are too many firms in low grades. This way of credit rating would not be practiced in real world applications. In general, commercial banks or other financial institutions want to grade the client firms according to the normal curve.²

² According to the internal policies of financial institutions, their respective credit rating distributions may differ. In this study, however, we assume that the normal distribution is

Fortunately, this problem may be solved through adjusting the class interval of DEA scores. Table 7 and Fig. 6 show the DEA score distribution of the client firms with modified class intervals.

As seen in Table 7 and Fig. 6, the modified distribution of the client firms guarantees that the bankruptcy ratio (% of Bad) decreases as the DEA score increases, and the frequency of the firms in each class (% of Sum) approximately follows the normal distribution.

5. CONCLUSIONS

This paper presents a new method of credit scoring using DEA. As opposed to broadly known multiple discriminant analysis, logistic regression analysis, and neural networks, DEA requires only ex post information to calculate credit scores. The discriminatory power of this method was also tested by comparing its results against those obtained by regression analysis and discriminant analysis, and by using actual bankruptcy cases. The empirical results suggest that this new approach can serve as a promising alternative for augmenting and/or replacing current credit scoring methods used by commercial banks and credit industry.

In terms of managerial implications, this method also gives a clear insight into how "bad" firms can improve their respective financial credibility. From the empirical results, it is shown that "good" firms have higher liquidity, lower bank loans, higher capital adequacy, and more balance between their equities and fixed assets. We also suggested a practical credit rating method using the estimated DEA scores derived by the method.

In addition, this new method *ipso-facto* allows the commercial banks or other financial institutions to monitor the exposures of their respective credit portfolios on an ongoing basis and to take preventive actions against the clients' defaults in an early stage.

ACKNOWLEDGEMENT

This work was supported by the Brain Korea 21 Project in 2004.

REFERENCES

- [1] Altman, E.I., "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, XXIII(4), pp589-609, 1968.
- [2] Anderson, T.W., *An Introduction to Multivariate Statistical Analysis*, New York, NY: Wiley, 1984.
- [3] Basel Committee on Banking Supervision, *Credit risk modeling: Current Practices and Applications*, Basle: Basel Committee Publications, 1999.
- [4] Beaver, W., "Financial ratios as predictors of failure," *Journal of Accounting Research*, pp71-102, Apr. 1966.
- [5] Bryant, S.M., "A Case-based Reasoning Approach to Bankruptcy Prediction Modeling," *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 6, No. 3, pp195-214, 1997.
- [6] Buta, P., "Mining for Financial Knowledge with CBR," *AI Expert*, Vol. 9, No. 2, pp34-41, 1994.
- [7] Charnes, A., W.W. Cooper, E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, Vol. 2, pp429-444, 1978.
- [8] Chen, M.C., S.H. Huang, "Credit Scoring and Rejected Instances Reassigning through Evolutionary Computation Techniques," *Expert Systems with Applications*, Vol. 24, pp433-441, 2003.
- [9] Cielen, A., K. Vanhoof, *Bankruptcy prediction using a data envelopment analysis*, Manuscript, Limburg University, Diebenpeek, 1999.
- [10] Coakley, J.R., C.E. Brown, C.E., "Artificial Neural Networks in Accounting and Finance: Modeling Issues," *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 9, No. 2, pp119-144, 2000.
- [11] Davis R.H., D.B. Edelman, A.J. Gammerman, "Machine Learning Algorithms for Credit-Card Applications," *IMA Journal of Mathematics Applied in Business and Industry*, Vol. 4, pp43-51, 1992.
- [12] Desai, V.S., J.N. Conway, G.A. Overstreet Jr., "Credit Scoring Models in the Credit Union Environment Using Neural Networks and Genetic Algorithms," *IMA Journal of Mathematics Applied in Business and Industry*, Vol. 8, pp324-346, 1997.
- [13] Desai, V.S., J.N. Crook, G.A. Overstreet Jr., "A Comparison of Neural Networks and Linear Scoring Models in the Credit Union Environment," *European Journal of Operational Research*, Vol. 95, pp24-37, 1996.
- [14] Dillon, W.R., M. Goldstein, *Multivariate Analysis Methods and Applications*, New York: Wiley, 1984.
- [15] Dimitras, A.I., R. Slowinski, R. Susmaga, C. Zopounidis, "Business failure prediction using rough sets," *European Journal of Operational Research*, Vol. 17, No. 3, pp263-280, 1999.
- [16] Dimitras, A.I., S.H. Zanakis, C. Zopounidis, C., "A Survey of Business Failure with an Emphasis on Prediction Methods and Industrial Applications," *European Journal of Operational Research*, Vol. 90, No. 3, pp487-513, 1996.
- [17] Eisenbeis, R.A., "Problems in applying discriminant analysis in credit scoring models," *Journal of Banking and Finance*, Vol. 2, pp205-219, 1978.
- [18] Emel, A.B., M. Oral, A. Reisman, R. Yolalan, "A Credit Scoring Approach for the Commercial Banking Sector," *Socio-Economic Planning Sciences*, Vol. 27, pp103-123, 2003.
- [19] English, W.B., W.R. Nelson, *Bank risk rating of business loans*, Board of Governors of the Federal Reserve System Finance and Economics Discussion Series, November, 1998.
- [20] Falbo, P., "Credit Scoring by Enlarged Discriminant Analysis," *OMEGA*, Vol. 19, No. 4, pp275-289, 1991.
- [21] Federal Reserve System Task Force on Internal Credit Risk Models, *Credit risk models at major US banking institutions: current state of the art and implications for assessments of capital adequacy*, Federal Reserve Bank Board of Governors, Supervisory Staff Reports, Washington, 1998.
- [22] Frydman H.E., E.I. Altman, D. Kao, "Introducing Recursive Partitioning for Financial Classification: the case of Financial Distress," *Journal of Finance*, Vol. 40, No. 1, pp269-291, 1985.
- [23] Hand, D.J. *Discrimination and Classification*, New York, NY: Wiley, 1981.
- [24] Jensen, H.L., "Using Neural Networks for Credit Scoring," *Managerial Finance*, Vol. 18, pp15-26, 1992.
- [25] Johnson, R.A., D.W. Wichern, *Applied Multivariate Statistical Analysis (Fourth Edition)*, Upper Saddle River, NJ: Prentice-Hall, 1998.
- [26] Lee, G., T.K. Sung, N. Chang, N., "Dynamics of Modeling in Data Mining: Interpretive Approach to Bankruptcy Prediction," *Journal of Management Information Systems*, Vol. 16, pp63-85, 1999.
- [27] Lee, H., H. Jo, I. Han, I., "Bankruptcy Prediction Using Case-based Reasoning, Neural Networks, and Discriminant Analysis," *Expert Systems With Applications*, Vol. 13, pp97-108, 1997.
- [28] Lee, T.S., C.C. Chiu, C.J. Lu, I.F. Chen, "Credit Scoring Using Hybrid Neural Discriminant Technique," *Expert Systems with Applications*, Vol. 23, pp245-254, 2002.
- [29] Lopez, J.A., M.R. Saldenberg, "Evaluating credit risk models," *The Journal of Banking and Finance*, Vol. 24, No. 1-2, pp151-65, 2000.
- [30] Malhotra, R., D.K. Malhotra, D.K., "Differentiating Between Good Credits and Bad Credits Using Neuro-fuzzy Systems," *European Journal of Operational Research*, Vol. 136, No. 2, pp190-211, 2002.
- [31] Martin, D., "Early Warning of Bank Failure: A Logit Regression Approach," *Journal of Banking and Finance*, Vol. 1, pp249-276, 1997.
- [32] Morrison, D.F., *Multivariate Statistical Methods*, New York, NY: McGraw-Hill, 1990.
- [33] Oral, M., R. Yolalan, "An empirical study on measuring operating efficiency and profitability of bank branches," *European Journal of Operational Research*, Vol. 46, pp282-294, 1990.
- [34] Peel, M.J., D.A. Peel, P.F. Pope, "Predicting corporate failure-some results for the UK corporate sector," *Omega*, Vol. 14, No. 1, pp5-12, 1986.
- [35] Reichert, A.K., C.C. Cho, G.M. Wagner, "An Examination of the Conceptual Issues Involved in Developing Credit-Scoring Models," *Journal of Business and Economic Statistics*, Vol. 1, pp101-114, 1983.
- [36] Roy, B., "The outranking approach and the foundations of ELECTRE methods," *Theory and Decision*, Vol. 31, pp49-73, 1991.
- [37] Simak, P.C., *DEA based analysis of corporate failure*, Manuscript, University of Toronto, Toronto, 1999.
- [38] Tam, K.Y., M.Y. Kiang, "Managerial Applications of Neural Networks: the Case of Bank Failure Predictions," *Management Science*, Vol. 38, No. 7, pp926-947, 1992.
- [39] Treacy, W.F., M. Carey, "Credit risk rating at large US banks," *The Journal of Banking and Finance*, Vol. 24, No. 1-2, pp167-201, 2000.
- [40] Trout, M.D., A. Rai, A. Zhang, "The potential use of DEA for credit applicant acceptance systems," *Computers and Operations Research*, Vol. 23, No. 4, pp405-408, 1996.
- [41] West, D., "Neural Network Credit Scoring Models," *Computers & Operations Research*, Vol. 27, pp1131-1152, 2000.
- [42] Yeh, Q.J., "The application of data envelopment analysis in conjunction with financial ratios for bank performance evaluation," *Journal of the Operational Research Society*, Vol. 47, pp980-988, 1996.
- [43] Zhang, G.P., "Neural Networks for Classification: A Survey," *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, Vol. 30, No. 4, pp451-462, 2000.
- [44] Zopounidis, C., M. Doumpos, "Developing a multicriteria decision support system for financial classification problems: the Finclas system," *Optimization Methods and Software*, Vol. 8, pp277-304, 1998.

adequate for the demonstration purpose.

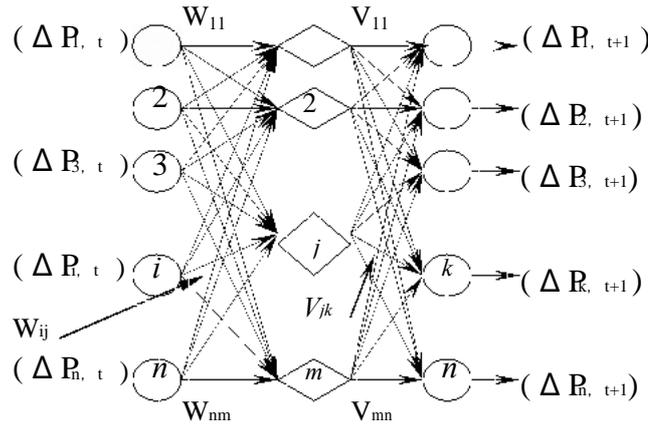


Figure1 : Capital Market Structure Model

In this model, there are n assets and m kinds of bargainer. Bargainer is partitioned into m kinds due to the degree of averseness to risk. The degree of averseness to risk of bargainer j is θ_j . In time t , the price of asset k is $P_{k,t}$, $k=1\sim n$, and its price fluctuating rate is shown in equation (1):

$$\Delta P_{k,t} = (P_{k,t} - P_{k,t-1}) / P_{k,t-1} \quad (1)$$

The weight of influence from price fluctuating rate $\Delta P_{k,t}$ to bargainer j is w_{kj} , the whole influence come from the price fluctuating rate of n asset of bargainer j is $\sum_{k=1}^n w_{kj} \Delta p_{k,t}$. deducting from θ_j , the real influence to the decision of bargainer j is $\sum_{k=1}^n w_{kj} \Delta p_{k,t} - \theta_j$. If

$\sum_{k=1}^n w_{kj} \Delta p_{k,t} - \theta_j \neq 0$, bargainer will changing his portfolio of assets. The influence of this changing to whole market will be $f(\sum_{k=1}^n w_{kj} \Delta p_{k,t} - \theta_j)$

The function $f(x)$ means the transformation from this influence to the changing of market. The weight of influence from bargainer j to asset k is v_{jk} , so the influence from the portfolio changing of bargainer j to the price fluctuating rate of asset k is $v_{jk} f(\sum_{k=1}^n w_{kj} \Delta p_{k,t} - \theta_j)$. All the influence from portfolio changing of m kind of bargainer to the price fluctuating rate of asset k is

$$\sum_{j=1}^m v_{jk} f(\sum_{k=1}^n w_{kj} \Delta p_{k,t} - \theta_j), \text{ which is shown in}$$

equation(2) :

$$\Delta P_{k,t+1} = \sum_{j=1}^m v_{jk} f(\sum_{k=1}^n w_{kj} \Delta p_{k,t} - \theta_j) \quad (2)$$

The price of asset k in time $t+1$ is shown in equation(3) :

$$P_{k,t+1} = P_{k,t} (1 + \Delta P_{k,t+1}) \quad (3)$$

In this model, the function $f(x)$ is the key factor in describing the process of price fluctuating of asset k deduced from $\Delta P_{k,t}$ to $\Delta P_{k,t+1}$.

According to the Prospect theory of Kahneman and Tversky^[2], the relationship between the subjective value and real surplus can be described by S curve. Because the behavior of bargainer is decided by his subjective value in bargaining, here we assume that there is a relationship which can be described as S curve between the influence of bargainer's asset portfolio changing to asset price and real surplus. It is showed as follow:

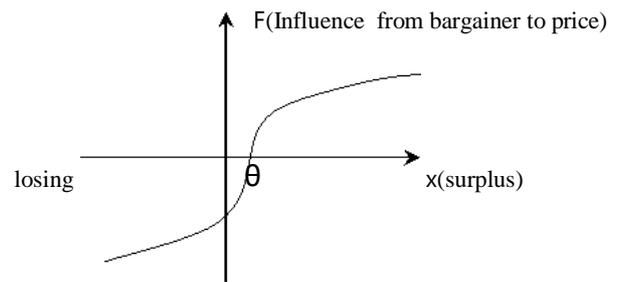


Figure 2 : Relationship between Influence from Bargainer to Price and Surplus

Because the sensitivity of every bargainer to losing is higher than which to earning, so here the slope of curve above axis X is less than which under axis X in figure 2.

Due to the Hypothesis, value of $f(x)$ is inside the zone $(-1, 1)$:

$$f(x) = \begin{cases} a \cdot b \cdot \tan \text{sig}(x), x \in (\theta, \theta + b) \\ 0, & x = \theta \\ b \cdot \tan \text{sig}(x), x \in (\theta - b, \theta) \end{cases} \quad (4)$$

In Chinese stock market, if daily price is applies, the constant b should be 10%.

To proof-test capability of the model, here the method of chaos forecast jointed to forecast price of stock in Shanghai Stock Market through the model.

To forecast under chaos theory, the historic time series data of a stock must be rebuilt- in special phase space, so the model must predigested. In the new model, the historic time series data is rebuilt- in a phase space with n dimensions.

data of dimension $K(k=1 \sim n-1)$ bargainer $j (j=1 \sim m)$ data of dimension n

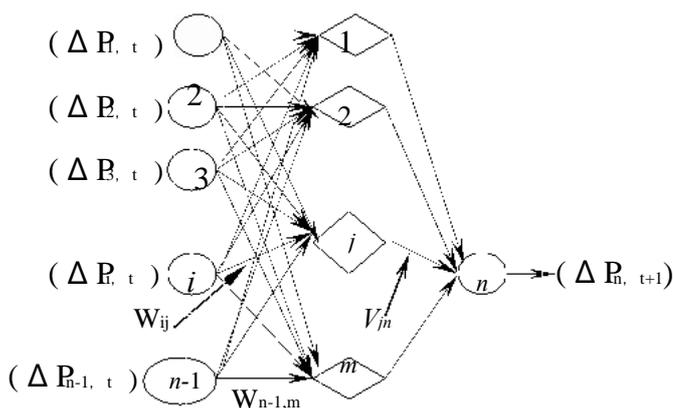


Figure 3 : Predigested CMSM

Here it is assumed that bargainer change their portfolios according data $\Delta P_{k,t}$, $k=1 \sim n-1$, influence of this changing to the price will be $\Delta P_{n,t}$

$$\Delta P_{n,t} = \sum_{j=1}^m v_{j,k} \sum_{k=1}^{n-1} (w_{kj} \Delta p_{k,t} - \theta_j) \quad (5)$$

there is no changing with function $f(x)$.

Weights w_{kj} , v_{jk} and averseness to risk θ_j can be accounted by error propagation.

2. A RESEARCH OF STOCK PRICE FORECAST

Some researches show that Shanghai Stock Market is a chaos economy^{[3][4]}, and those papers also show that it is available^[5] in short-term forecast. Here the historic daily price fluctuating of stock 600622 from May17,1993 to

Oct.10,2003 are apply to this research, software MATLAB 6.01 is also applies in data computing.

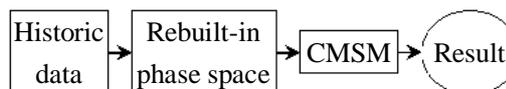


Figure 4 : Process of Chaos Forecast through CMSM

First of all, the length of non-periodic cycle of stock price fluctuating should be accounted by V-Stat. developed by Hurst:

$$V_n = (R/S)_n / \sqrt{n} \quad (6)$$

$(R/S)_n$ can be accounted as follows :

$$x_m = (x_1 + \dots + x_n) / n, \quad x (r = 1 \sim n) \quad (7)$$

$$\Delta x_i = (x_i - x_m), \quad i = 1 \sim n \quad (8)$$

$$y_r = \sum_{i=1}^r \Delta x_i; \quad r = 1 \sim n \quad (9)$$

$$R_n = \max(y_r) - \min(y_r); \quad r = 1 \sim n \quad (10)$$

$$S_n = \sqrt{\frac{1}{n} \sum_{r=1}^n (x_r - x_m)^2} \quad (11)$$

$$R_n / S_n = \frac{\max(y_r) - \min(y_r)}{\sqrt{\frac{1}{n} \sum_{r=1}^n (x_r - x_m)^2}} \quad (12)$$

If $(R/S)_n$ is independence and stochastic, and changing with the square root of time, we will get a horizontal line in a figure which with the X axis of $\text{Log}(n)$ and the Y axis of $(R/S)_n$. If the changing rate of $(R/S)_n$ is faster than which of time square root, we will get a decline curve in the figure. Otherwise, we will get an upswept curve.

When the curve changing its trend, it means the end of non-periodic cycle, so the length of non-periodic cycle is gotten.

When analyzing the historic price fluctuating of stock 600622, we find that the non-periodic cycle is 30 days. It means that the available length in forecasting is 30 days.

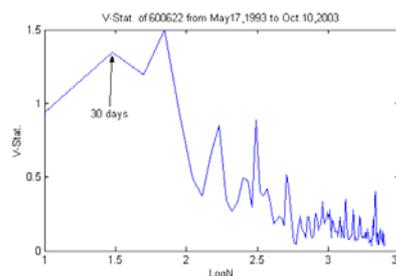


Figure 5 : V-Stat. of 600622

According to Wolf.,^[6] the product of the number of phase space dimension and arrearage equal to the length of non-periodic cycle. There are 5 working days every week, so the number of dimension is 5, Accordingly, the arrearage should be 6 days.

After the process of rebuilt in phase space, the one dimension historic time series is convert into 5 dimension time series. 4 of 5 dimension time series is taken as the input data, and the 5th dimension of time series is taken as the output data. When forecasting, we can use the data of 1st and 7th and 13th and 19th to predict the value of 25th. According to the principle of pricing, it can be explained that bargainer change their stock portfolios due to price fluctuating occurred at 1st and 7th and 13th and 19th those 4 Days, all influences of portfolio changing will decide the pricing at the day of 25th. Most of historic data are apply to train CMSM, less of them are apply to proof-test the accuracy of the CMSM. The constant of a is 0.816, and the number of kinds of bargainer is 9. Percentage of daily price fluctuating forecasted and its error list as follows:

Table 1: Percent of daily price fluctuating forecasted and pricing error

NO.	Real value (%)	Accounted value(%)	Error(%) of pricing
1	0.13	0.84	0.71
2	-0.41	2.68	3.09
3	-0.88	2.02	2.90
4	1.74	-0.94	-2.68
5	-0.32	1.42	1.75
6	-1.06	0.33	1.38
7	-0.96	0.85	1.80
8	-3.04	3.24	6.28
9	-1.01	-3.10	-2.09

3. CONCLUSION

Through the work of test-proof about the CMSM in forecasting the daily price fluctuating, we can find that the CMSM is applies to price forecasting in chaos economy because of the accuracy which shows that 8 of the 9 cases in price forecasting with its error less than 3.5% and there is only one case of price forecasting with its error more than 5%. So CMSM is apply to consult in decision-making.

REFERENCES

- [1] YANGYi-wen,LIUGui-zhong,ZHANGZong-ping, Stock market trend prediction based on neural networks, multiresolution analysis and dynamical reconstruction, Systems Engineering :Theory & Practice, No. 8, pp19-23, 2001.
- [2] Andrei.Shleifer, Inefficient Markets: An introduction to behavioral finance, Oxford University Press, 2000, P120-150.
- [3] Zhang Yongdong, An examination of nonlinearity and chaos in the Shanghai stock market, Journal of Engineering management, No. 3, pp21-26, 2003.
- [4] Gao Hongbing, Pan Jin, Chen Hongmin, Judgement of chaos in our securities market, System Engineering, No. 11, pp28-32, 2000.
- [5] GUO Gang, SHI Zhong-ke, DAI Guan-zhong, Stock market multi-step forecast using chaos theory, Information and Control, No. 4, pp77-81, 2000.
- [6] Wolf, A., Swift, J.B., Swinney, H.L., and Vastano, J.A. "Determining Lyapunov Exponents from a Time Series," Physic 16D, July 1985.

A Three-phased Online Association Rule Mining Approach for Diverse Mining Requests

Chingyao Wang¹, Shianshyong Tseng¹, Tzungpei Hong², Yianshu Chu¹

¹ Institute of Computer and Information Science, National Chiao-Tung University,
Hsinchu, 300, Taiwan, China

² Department of Electrical Engineering, National University of Kaohsiung,
Kaohsiung, 811, Taiwan, China
tphong@nuk.edu.tw, {cywang, sstseng, yschu}@cis.nctu.edu.tw

ABSTRACT

In the past, most incremental mining and online mining algorithms considered finding the set of association rules or patterns consistent with the entire set of data inserted so far. Users can not easily obtain the results from their only interested portion of data. For providing ad-hoc, query-driven and online mining supports, we first propose a relation called *multidimensional pattern relation* to structurally and systematically store the context information and the mining information for later analysis. Each tuple in the relation comes from an inserted dataset in the database. This concept is similar to the construction of a data warehouse for OLAP. However, unlike the summarized information of fact attributes in a data warehouse, the mined patterns in the multidimensional pattern relation can not be directly aggregated to satisfy users' mining requests. We then develop an online mining approach called *Three-phased Online Association Rule Mining* (TOARM) based on the proposed multidimensional pattern relation to support online generation of association rules under multidimensional considerations. Experiments for both homogeneous and heterogeneous datasets are made, with results showing the effectiveness of the proposed approach.

Keywords: association rule, incremental mining, multidimensional mining, constraint-based mining, data warehouse

1. INTRODUCTION

Data mining technology has become increasingly important in the field of large databases and data warehouses. This technology helps discover non-trivial, implicit, previously unknown and potentially useful knowledge [3][9][16], thus being able to aid managers in making good decision. Among various types of databases and mined knowledge, mining association rules from transaction databases is the most interesting and popular. Previous works on mining association rules could be classified into batch mining approaches [2][4][5][7][20][22][24] and incremental mining approaches [10][11][13][18][23][25] according to the processing ways. Most of them have focused on finding association rules or patterns in a specified part of a database [15]. Some contexts (circumstance information) such as region, time and branch have usually been ignored in mining requests. Users can not easily obtain association rules or patterns from their only interested portion of data. However, decision-makers usually diversely consider problems at different aspects [14][15][16]. They may need to analyze market demands, customer preferences, localities, and short-term/long-term trends. They may also want to understand the change of discovered patterns or rules in different dimensions. This may decrease the usage of mining in online decision support for multidimensional data.

In this paper, we attempt to extend the concept of *effectively utilizing previously discovered patterns* in incremental data mining to support online generation of

association rules under multidimensional considerations. We first propose the *multidimensional pattern relation* to structurally and systematically store the additional context information and mining information for each inserted dataset. It is conceptually similar to the construction of a data warehouse for OLAP [8][19][26]. Both of them preprocess the underlying data in advance, integrate related information, and store the results in a centralized structural repository for later use and analysis. However, unlike the summarized information of fact attributes in a data warehouse, the mined patterns in the multidimensional pattern relation can not be directly aggregated to satisfy users' mining requests. We then develop a *Three-phased Online Association Rule Mining* (TOARM) approach to effectively and efficiently satisfy diverse mining requests. It mainly consists of three phases, *generation of candidate itemsets*, *reduction of candidate itemsets*, and *generation of association rules*. The phase for *generation of candidate itemsets* selects the tuples satisfying the context constraints in a mining request and generates the candidate itemsets from the matched tuples. After that, the phase for *reduction of candidate itemsets* calculates the upper-bound supports of the candidate itemsets and adopts two pruning strategies to reduce the number of candidate itemsets. Finally, the phase for *generation of association rules* finds the final large itemsets and then derives the association rules from them. Experimental results also show the effectiveness of the proposed TOARM approach.

2. RELATED WORK

An association rule indicates a relationship among items such that the occurrence of certain items in a transaction would imply the occurrence of some other items in the same transaction. The process of mining association rules can roughly be decomposed into two tasks [4]: *finding large itemsets* and *generating interesting association rules*. The first task discovers the itemsets that satisfy a user-specified *minimum support* from a given database. It is used to obtain the statistically significant patterns. The second task finds the association rules that satisfy a user-specified *minimum confidence* from the large itemsets. Since this process is rather costly and time-consuming, some famous mining algorithms, such as Apriori [4], DIC [7], DHP [22], Partition [24], Sampling [20] and GSP [5], were proposed to achieve this purpose. Among them, the Apriori algorithm, which is the best well-known, utilizes a level-wise candidate generation approach to reduce its search space, such that only the large itemsets found in the previous level are treated as seeds for generating the candidate itemsets in the current level. This level-by-level property can greatly reduce the number of itemsets to be considered in a mining process. Many following algorithms were then based on this property and attempted to further reduce candidate itemsets and I/O costs. Comprehensive overviews can be referred to in [9][16].

Most of the mining algorithms process data in a batch way and must re-process the entire database whenever either the data stored in a database or the thresholds (i.e. the minimum support or the minimum confidence) set by users are changed. They do not utilize previously mined patterns for later maintenance, and may require considerable computation time to obtain the updated set of association rules or patterns [10]. Recently, some researchers have developed incremental mining algorithms to maintain association rules without re-processing the entire database whenever the database is updated. Examples include the FUP-based algorithms proposed by Cheung et al. [10][11], the adaptive algorithm proposed by Sarda and Srinivas [23], the incremental mining algorithm based on the concept of *pre-large itemsets* proposed by Hong et al. [18], and the incremental updating technique based on the concept of *negative border* proposed by Thomas et al. [25] and Feldman et al. [13]. The common idea of the above researches lies in that the previously mined patterns are stored in advance for later usage. When new transactions are inserted or old records are deleted, a large part of the final results can be obtained by comparing the patterns mined from the newly inserted transactions or deleted records with the pre-stored mined knowledge. Only a small portion of patterns needs to be re-processed against the entire database. Much computation time can thus be saved in this way. Among the above approaches, the FUP-based algorithms [10][11] store the previously mined large

itemsets for later maintenance. Some other approaches utilize the *pre-large itemsets* [18] and the *negative border* [13][25] to enlarge the amount of pre-stored mined information for further improving the maintenance performance at the expense of storage spaces.

3. THE MULTIDIMENSIONAL PATTERN RELATION

A multidimensional pattern relation schema *MPR* is a special relation schema for storing mining information. An *MPR* consists of three types of attributes, *identification (ID)*, *context*, and *content*. There is only one identification attribute for an *MPR*. It is used to uniquely label the tuples. Context attributes describe the contexts (circumstance information) of an individual block of data which are gathered together from a specific business viewpoint. Examples of context attributes are region, time and branch. Content attributes describe available mining information which is discovered from each individual block of data by a batch mining algorithm. Examples of content attributes include the number of transactions, the number of mined patterns, and the set of previously mined large itemsets with their supports.

The set of all previously mined patterns with their supports for an individual block of data is called a *pattern set (ps)* in this paper. Assume the minimum support is s and there are l large itemsets discovered from an individual block of data. A pattern set can be represented as $ps = \{(x_i, s_i) \mid s_i \geq s \text{ and } 1 \leq i \leq l\}$, where x_i is a large itemset and s_i is its support. The pattern set is thus a principal content attribute for an inserted block of data.

A multidimensional pattern relation schema *MPR* with n_1 context attributes and n_2 content attributes can be represented as $MPR(ID, CX_1, CX_2, \dots, CX_{n_1}, CN_1, CN_2, \dots, CN_{n_2})$, where *ID* is an identification attribute, $CX_i, 1 \leq i \leq n_1$, is a context attribute, and $CN_i, 1 \leq i \leq n_2$, is a content attribute. Assume a multidimensional pattern relation *mpr*, which is an instance of the given *MPR*, includes tuples $\{t_1, \dots, t_m\}$. Each tuple $t_i = (id_i, cx_{i1}, cx_{i2}, \dots, cx_{in_1}, cn_{i1}, cn_{i2}, \dots, cn_{in_2})$ in *mpr* indicates that for the block of data under the contexts of cx_{i1}, cx_{i2}, \dots , and cx_{in_1} , the mining information contains cn_{i1}, cn_{i2}, \dots , and cn_{in_2} .

Example 1: Table 1 shows a multidimensional pattern relation with the initial minimum support set at 5%. *ID* is an identification attribute, *Region*, *Branch* and *Time* are context attributes, and *No_Trans*, *No_Patterns* and *Pattern_Sets* are content attributes. The *Pattern_Sets*

attribute records the sets of mined large itemsets from the previous data blocks. For example, the tuple with $ID = 1$ shows that seven large itemsets, $\{(A, 10\%), (B, 11\%), (C, 9\%), (AB, 8\%), (AC, 7\%), (BC, 6\%), (ABC, 6\%)\}$, are discovered from 10000 transactions and under the contexts of $Region = CA$, $Branch = San Francisco$ and $Time = 2003/10$. The other tuples have similar meaning. \clubsuit

Table 1: A multidimensional pattern relation with minimum support = 5%

ID	Region	Branch	Time	No_ Trans.	No_ Patterns	Pattern_Sets (Itemset, Support)
1	CA	San Francisco	2003/10	10000	7	(A,10%),(B,11%),(C,9%),(AB,8%),(AC,7%),(BC,6%),(ABC,6%)
2	CA	San Francisco	2003/11	15000	3	(A,5%),(B,7%),(C,5%)
3	CA	San Francisco	2003/12	12000	2	(A,5%),(C,9%)
4	CA	Los Angeles	2003/10	20000	4	(A,8%),(B,6%),(C,7%),(AC,6%)
5	CA	Los Angeles	2003/11	25000	2	(A,5%),(C,6%)
6	CA	Los Angeles	2003/12	30000	4	(A,6%),(B,6%),(C,9%),(AB,6%)
7	NY	New York	2003/10	18000	3	(B,8%),(C,7%),(B,C,6%)
8	NY	New York	2003/11	18500	2	(B,8%),(C,6%)
9	NY	New York	2003/12	19000	5	(A,5%),(B,9%),(C,8%),(D,6%),(BC,6%)

4. MULTIDIMENSIONAL ONLINE MINING FOR ASSOCIATION RULES

The goal of online mining is to find the association rules satisfying the constraints in a mining request on line. The types of mining requests allowed can grow up through the usage of the proposed multidimensional pattern relation. In this paper, an online mining approach called *Three-phased Online Association Rule Mining* (TOARM) is proposed to achieve the mining task from a multidimensional pattern relation. TOARM first selects the tuples from the relation satisfying the constraints in a mining request. It then integrates and outputs the mining information in these tuples to users. Before describing the TOARM approach, we first formally define the problem to be solved and some related terminology. Some lemmas are also derived (The detailed proofs are omitted here).

Assume $mpr = \{t_1, t_2, \dots, t_n\}$ is a multidimensional pattern relation based on an initial minimum support s . Given a mining request q with a set of contexts cx_q , a new minimum support s_q ($s_q \geq s$), and a new minimum confidence $conf_q$, the proposed algorithm will effectively and efficiently derive the association rules satisfying s_q , $conf_q$ and cx_q . A tuple with cx in a multidimensional pattern relation is called a *matched*

tuple. Let t_i denote the i -th tuple in a multidimensional pattern relation, $t_i.trans$ denote the number of transactions kept in t_i , $t_i.ps$ denote the pattern set in t_i , and $t_i.s_x$ denote the actual support of an itemset x in t_i .

Lemma 1: For each itemset x satisfying s_q and cx_q in a mining request q , there exists at least a matched tuple t , such that $t.s_x$ satisfies s_q . \clubsuit

Lemma 2: For each itemset x satisfying s_q and cx_q in a mining request q , it must be among the candidate itemsets obtained by collecting the ones whose supports are larger than or equal to s_q in at least one matched tuple. \clubsuit

Lemma 3: If x is a candidate itemset, then $\forall x' \subset x, x'$ is also a candidate itemset. \clubsuit

The *appearing count* $Count_x^{appearing}$ of a candidate itemset x is defined as the count of x calculated from the matched tuples in which x appears. Thus:

$$Count_x^{appearing} = \sum_{t_i \in \text{matched tuples} \ \& \ x \in t_i.ps} t_i.trans * t_i.s_x. \quad (1)$$

The *upper-bound count* $Count_x^{UB}$ of a candidate itemset x is defined as the upper bound count of x calculated from the matched tuples in which x does not appear. Thus:

$$Count_x^{UB} = \sum_{t_i \in \text{matched tuples} \ \& \ x \notin t_i.ps} (t_i.trans * s - 1). \quad (2)$$

Let $Match_Trans$ denote the number of transactions in the matched tuples. Thus:

$$Match_Trans = \sum_{t_i \in \text{matched tuples}} t_i.trans. \quad (3)$$

The *upper-bound support* s_x^{UB} of a candidate itemset x is thus calculated as:

$$s_x^{UB} = \frac{Count_x^{appearing} + Count_x^{UB}}{Match_Trans}. \quad (4)$$

Lemma 4: If x is a candidate itemset and s_x is its actual support, then $s_x \leq s_x^{UB}$. \clubsuit

Lemma 5: If x is a candidate itemset, then $\forall x' \subset x, s_{x'}^{UB} \geq s_x^{UB}$. \clubsuit

Lemma 6: If a candidate itemset x is contained in all the matched tuples, then $s_x^{UB} = s_x$. \clubsuit

The Three-phased Online Association Rule Mining (TOARM) approach:

INPUT: A multidimensional pattern relation based on an initial minimum support s and a mining request q with a set of contexts cx_q , a minimum support s_q and a minimum confidence $conf_q$.

OUTPUT: A set of association rules satisfying the mining request q .

Phase 1: Generation of candidate itemsets:

(a) Select the tuples satisfying cx_q from the multidimensional pattern relation.

(b) Gather the candidate itemsets appearing in the matched tuples.

(c) Calculate $Count_x^{appearing}$ and $Count_x^{UB}$ for each candidate itemset x .

Phase 2: Reduction of candidate itemsets:

(a) Calculate the upper-bound support s_x^{UB} of each candidate itemset x by the formula:

$$s_x^{UB} = \frac{Count_x^{appearing} + Count_x^{UB}}{Match_Trans}$$

(b) Discard the candidate itemset x and its proper supersets from the candidate set if $s_x^{UB} \leq \xi$.

(c) Put x into the set of large itemsets if $s_x^{UB} = \frac{Count_x^{appearing}}{Match_Trans}$ and $s_x^{UB} \geq \xi$.

Phase 3: Generation of association rules:

(a) Check whether each remaining candidate itemset x is large by scanning the underlying blocks of data for the matched tuples in which x does not appear.

(b) Generate the association rules satisfying the minimum confidence $conf_q$ from the set of large itemsets.

The TOARM approach only considers the itemsets appearing in the matched tuples and satisfying the minimum support as the candidate ones. It also uses two pruning strategies to reduce the number of candidate itemsets. It therefore only needs to re-process the remaining candidate itemsets against the underlying blocks of data for the matched tuples in which they do not appear. Due to the above consideration, the cost of re-processing underlying blocks of data by the TOARM approach is less than that by typical batch mining or incremental mining approaches.

Example 2: For the multidimensional pattern relation given in Table 1, assume a mining request q is to get the patterns under the contexts cx of $Region = CA$ and $Time = 2003/11 \sim 2003/12$ and satisfying the minimum support $\xi = 5.5\%$. According to Lemma 2, the set of candidate itemsets is $\{\{A\}, \{B\}, \{C\}, \{AB\}\}$, which is the union of the itemsets appearing in the pattern sets and with their supports larger than 5.5%. Among these candidate itemsets, in Phase 2, the TOARM approach can remove the candidate itemsets $\{A\}$ and $\{AB\}$ according to Lemmas 4 and 5, and put the candidate itemset $\{C\}$ into the set of large itemsets for q according

to Lemma 6. Only the remaining candidate itemset $\{B\}$ needs to be further processed in Phase 3.

6. EXPERIMENTS

The experiments were implemented in Java on a workstation with dual XEON 2.8GHz processors and 2048MB main memory, running RedHat 9.0 operation system. The datasets were generated by a generator similar to that used in [4]. The generator first generated L maximal potentially large itemsets, each with an average size of I items. The items in a potentially large itemset were randomly chosen from the total N items according to its actual size. The generator then generated D transactions, each with an average size of T items. The items in a transaction were generated according to the L maximal potentially large itemsets in a probabilistic way.

The two groups of datasets generated in the above way and used in our experiments are listed in Table 2, where the datasets in the same group had the same D , T and I values but different L or N values. Each dataset was treated as a block of data in the database. Among the two groups, Group 2 could be thought of as heterogeneous because of its varied N values. This group of datasets was used to show the effect of heterogeneous blocks of data on our approach.

Table 2: The two groups of datasets generated for the experiments

Group	Size	Datasets	D	T	I	L	N
1	10	T10I8D10KL ^{1 to 10} T10I8D10KL ¹⁰	10000	10	8	200 to 245	100
2	10	T10I8D10KN ^{1 to 10} T10I8D10KN ¹⁰	10000	10	8	200	100 to 145

The TOARM and the Apriori algorithms were then run for Groups 1 and 2 along with different minimum supports ranging from 0.022 to 0.04 in the mining requests. The execution times spent by the two algorithms for each group are respectively shown in Figures 1 and 2. From Figures 1, it is easily seen that the execution time by the TOARM algorithm on Groups 1 was always much less than that by the Apriori algorithm. This is because the datasets in this group was homogeneous, meaning they used the same set of items in each group. In this situation, the number of candidate itemsets considered by the TOARM algorithm was much closer to the number of the final large itemsets than that by the Apriori algorithm. The former thus had a more compact candidate set than the latter.

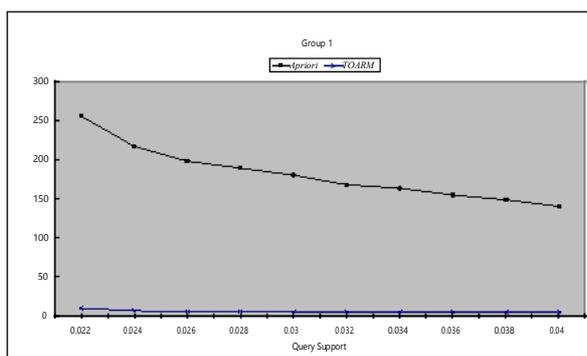


Figure 1: The execution time spent by the two algorithms for Group 1

On the contrary, the datasets in Group 2 were heterogeneous, meaning they used different sets of items. In this situation, the number of candidate itemsets considered by the TOARM algorithm was much larger than the number of the final large itemsets since most of the candidate itemsets appeared in only one or few tuples in the multidimensional pattern relation. But, since the TOARM algorithm adopted two pruning strategies in Phase 2 and only re-processed the remaining candidate itemsets in Phase 3 against the underlying datasets in which they do not appear, the execution time spent by the TOARM algorithm was usually still less than that spent by the Apriori algorithm. This is also consistent with the results shown in Figure 2.

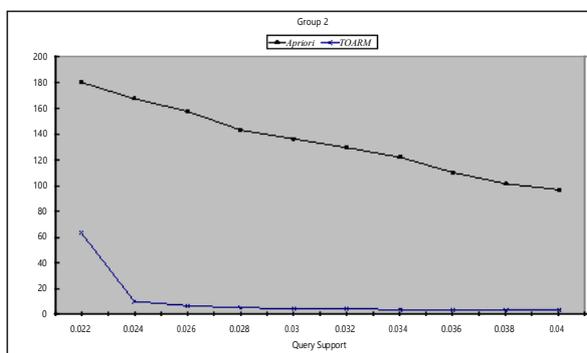


Figure 2: The execution time spent by the two algorithms for Group 2

7. CONCLUSION

In this paper, we have extended the concept of effectively utilizing previously discovered patterns in incremental mining to online decision support under multidimensional considerations. By structurally and systematically storing the additional context information and mining information in the multidimensional pattern relation, our proposed TOARM approach can easily and efficiently derive the association rules satisfying diverse user-concerned constraints. From the experimental results, the proposed TOARM approach is more efficient than the well-known Apriori approach especially for homogeneous datasets.

ACKNOWLEDGEMENT

This research was supported by the National Science Council of Taiwan, China under Grand No. NSC93-2752-E-009-006-PAE.

REFERENCES

- [1] C. C. Aggarwal and P. S. Yu, "A New Approach to Online Generation of Association Rules," IEEE Transactions on Knowledge and Data Engineering, Vol. 13, No. 4, pp. 527-540, 2001.
- [2] R. Agrawal, T. Imielinski and A. Swami, "Mining Association Rules between Sets of Items in Large Database," ACM SIGMOD Conference, pp. 207-216, Washington DC, USA, 1993.
- [3] R. Agrawal, T. Imielinski and A. Swami, "Database Mining: A Performance Perspective," IEEE Transactions on Knowledge and Data Engineering, Vol. 5, No. 6, pp. 914-925, 1993.
- [4] R. Agrawal and R. Srikant, "Fast Algorithm for Mining Association Rules," ACM International Conference on Very Large Data Bases, pp. 487-499, 1994.
- [5] R. Agrawal and R. Srikant, "Mining Sequential Patterns," IEEE International Conference on Data Engineering, pp. 3-14, 1995.
- [6] K. Beyer and R. Ramakrishnan, "Bottom-Up Computation of Sparse and Iceberg CUBEs," ACM SIGMOD Conference, pp. 359-370, 1999.
- [7] S. Brin, R. Motwani, J. D. Ullman and S. Tsur, "Dynamic Itemset Counting and Implication Rules for Market Basket Data," ACM SIGMOD Conference, pp. 255-264, Tucson, Arizona, USA, 1997.
- [8] S. Chaudhuri and U. Dayal, "An Overview of Data Warehousing and OLAP Technology," ACM SIGMOD Record, 26:65-74, 1997.
- [9] M. S. Chen, J. Han and P. S. Yu, "Data mining: An Overview from A Database Perspective," IEEE Transactions on Knowledge and Data Engineering, Vol. 8, No. 6, pp. 866-883, 1996.
- [10] D. W. Cheung, J. Han, V. T. Ng and C. Y. Wong, "Maintenance of Discovered Association Rules in Large Databases: An Incremental Updating Approach," IEEE International Conference on Data Engineering, pp. 106-114, 1996.
- [11] D. W. Cheung, S. D. Lee, and B. Kao, "A General Incremental Technique for Maintaining Discovered Association Rules," In Proceedings of Database Systems for Advanced Applications, pp. 185-194, Melbourne, Australia, 1997.
- [12] M. Fang, N. Shivakumar, H. Garcia-Molina, R. Motwani, J. D. Ullman, "Computing Iceberg Queries Efficiently," ACM International Conference on Very Large Data Bases, pp. 299-310, 1998.
- [13] R. Feldman, Y. Aumann, A. Amir, and H. Mannila, "Efficient Algorithms for Discovering Frequent Sets in Incremental Databases," ACM SIGMOD Workshop on DMKD, pp. 59-66, USA, 1997.

- [14] G. Grahne, L. V. S. Lakshmanan, X. Wang and M. H. Xie, "On Dual Mining: From Patterns to Circumstances, and Back," IEEE International Conference on Data Engineering, pp. 195-204, 2001.
- [15] J. Han, L. V. S. Lakshmanan and R. Ng, "Constraint-based, Multidimensional Data Mining," IEEE Computer Magazine, pp.2-6, 1999.
- [16] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2001.
- [17] C. Hidber, "Online Association Rule Mining," ACM SIGMOD Conference, pp. 145-156, USA, 1999.
- [18] T. P. Hong, C. Y. Wang and Y. H. Tao, "A New Incremental Data Mining Algorithm Using Pre-large Itemsets," International Journal on Intelligent Data Analysis, 2001.
- [19] W. H. Immon, *Building the Data Warehouse*, Wiley Computer Publishing, 1996.
- [20] H. Mannila, H. Toivonen and A.I. Verkamo, "Efficient Algorithm for Discovering Association Rules," The AAAI Workshop on Knowledge Discovery in Databases, pp. 181-192, 1994.
- [21] H. Mannila and H. Toivonen, "On an Algorithm for Finding all Interesting Sentences," The European Meeting on Cybernetics and Systems Research, Vol. II, 1996.
- [22] J. S. Park, M. S. Chen and P. S. Yu, "Using a Hash-based Method with Transaction Trimming for Mining Association Rules," IEEE Transactions on Knowledge and Data Engineering, Vol. 9, No. 5, pp. 812-825, 1997.
- [23] N. L. Sarda and N. V. Srinivas, "An Adaptive Algorithm for Incremental Mining of Association Rules," IEEE International Workshop on Database and Expert Systems, pp. 240-245, 1998.
- [24] A. Savasere, E. Omiecinski and S. Navathe, "An Efficient Algorithm for Mining Association Rules in Large Database," ACM International Conference on Very Large Data Bases, pp. 432-444, 1995.
- [25] S. Thomas, S. Bodagala, K. Alsabti and S. Ranka "An Efficient Algorithm for the Incremental Update of Association Rules in Large Databases," The International Conference on Knowledge Discovery and Data Mining, pp. 263-266, 1997.
- [26] J. Widom, "Research Problems in Data Warehousing," ACM International Conference on Information and Knowledge Management, 1995.

A User Preference Classification Method in Information Recommendation System

Chentung Chen¹, Weishen Tai²

¹ Department of Information Management, Da-Yeh University, Taiwan, China

² Institute of Management, School of Management, Da-Yeh University, Taiwan, China

^{1,2} {chtung, wstai}@mail.dyu.edu.tw

ABSTRACT

As information overload problem more serious on the Internet, it becomes an important issue for users to retrieve information effectively. An information recommendation system is helpful for providing user information meet he/she requirements appropriately. However the traditional recommendation concepts usual classify a user into one preference class. It seems unreasonable because a user may possess interests about many information classes generally. In this study we propose a new recommendation concept in information recommendation system, namely club member, differs from content-based and collaboration filter method. It can classify a user into some clubs which he/she interests with different preference degrees.

In order to classify users into multi-club with different preference degrees, fuzzy association rule based on data mining technology is applied in this study. Fuzzy association rules are generated by discovering and analyzing the members' feature in the same club. According to fuzzy association rules, a new user can be classified into some clubs that he/she may interest with preference degrees. It is helpful for an information recommendation system to provide user more suitable information in accordance with their preferences precisely.

Keywords: Preference classification, Recommendation system, Data Mining, Fuzzy association rule

1. INTRODUCTION

The Internet, with continuous progress in computer technology and the pervasion of network environments, has become a major channel for modern people to acquire information (Chen and Tai, 2003). As amounts of information increase on the Internet, Users also face the serious problem of information overload simultaneously (Ribak et al., 2003; Sugiyama et al., 2004; Fang and Sheng, 2004). Under this circumstance, Web tools have become more important for developing better solutions to assist people in acquiring information easily and effectively (Sugiyama et al., 2004).

Thus far, Recommendation Systems (RSs) are regarded as one of effective solutions for assisting people to acquire information on the Internet (Garcia et al., 2004). Furthermore, RSs differ from traditional search engine tools or portal websites; it applies active information recommendation methods to replace passive keyword search methods. It is helpful for user to acquire suitable information in less time and effort (Chen and Tai, 2003). Nowadays, content-based and collaborative filter are two popular means applied in the RSs. Although they really emerge excellent performance in information recommendation, there are some problems still cannot be resolve. In order to provide user more abundance and suitable information, a new recommendation concept is proposed in this study, namely club member, differs from content-based and collaboration filter method. It allows a user can acquire a variety of information from some clubs that he/she interests.

In order to classify users into multi-club with different preference degrees, fuzzy association rule based on data mining technology is applied in this study. Fuzzy association rules are generated by discovering and analyzing the members' feature in the same club. Those fuzzy association rules can classify a new coming user into those clubs that he/she maybe interests with preference degrees. It is helpful for an information RS to provide user more suitable information in accordance with their preferences precisely.

2. RECOMMANDATION SYSTEM

Generally speaking, RS is applied to recommend user suitable information, e.g. news, products or services, which meet he/she interests or requirements (Chen and Tai, 2003). For the sake of achieving this goal, it must rely on several effective methods to analyze or predict user's preference and purchase trend to improve the accuracy ratio of recommendation (Garcia et al., 2004; Zhang and Iyenger, 2002). Recently, most of recommendation means can be divided into collaborative filter and content-based. Their definitions, advantages and drawbacks are described as follows (Weng and Liu, 2004; Wang et al., 2004)

2.1. Content-based approach

It recommends user information according to the results of similarity measurement between item-feature and target-feature. In general, a term vector is applied to represent information content and relevant profile (e.g. category, user and information) (Shih and Karger, 2004;

Wang et al., 2004). A recommendation system will process similarity measurement and determine how to recommend information to user (as shown in Fig. 1).

There are several ways can be applied as follows:

1. According to the result of similarity measurement between each piece of information and each category, information can be classified to appropriate category. A user can acquire recommendation information from his/her preference categories.
2. According to the result of similarity measurement between each piece of information and each user profile, information will be recommended to those users whose feature is similar to information feature directly.
3. According to the result of similarity measurement between each piece of information and the other information, information will be recommended to those users who had read relevant information in the past.

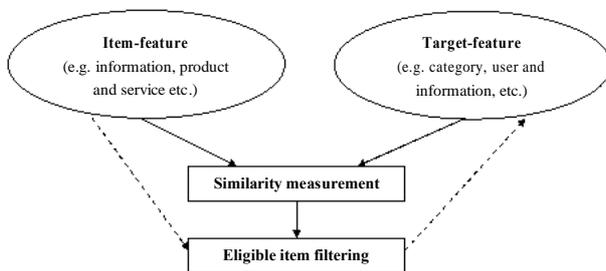


Fig. 1 Concept of content-based approach

Contented-based approach possesses advantages include:

1. A user will acquire information meet his preference feature or category appropriately.
2. Even though a piece of information cannot interest anyone in the past, it is still possible to be recommended to someone.
3. When user accepts recommendation information and makes feedback to the system, collaborative filter can support the RS (Weng and Liu, 2004).

It also possesses drawbacks include:

1. Multi-media information, such as image, sound and video data cannot be transformed into term vector. It makes the RS cannot define the content of those data type and process similarity measurement.
2. If the RS recommends user information according to user past browsing record or purchasing history, unique or different information will hardly be recommended to user (Weng and Liu, 2004).
3. Both the content of term vector and the number of term will determine the similarity degree between two features.
4. A user cannot acquire any information come from the other user with similar interests.

2.2. Collaborative filter analysis

In this concept, a preliminary supposition is "a set of users possess similar feature (e.g. preference, taste, purchasing trend and browsing pattern) will possess the

same interests. So a user will interest these information comes from the other members' past browsing record or purchasing history in the same group." Therefore, users are clustered to several groups according to their feature. The RS will recommend information to all members in the same group (as shown in Fig. 2). Collaborative filter analysis possesses advantages include:

1. The RS recommends information to all members in the same group; meanwhile, item-feature need not be analyzed in advance. This advantage makes multi-media information can be recommended without content definition or similarity measurement.
2. The recommended item could be different from users' past preference or experiments. Therefore, it is possible to draw up potential needs and interests of users (Herlocker et al., 2000).
3. Even though a user has not any browsing record or purchasing history, he/she can still acquire information from the other members in the same group.

It also possesses drawbacks include:

1. The RS must perform user clustering to locate user-group with similar interest or pattern.
2. Information (item) not yet rated or purchased could not be recommended to other users (Weng and Liu, 2004).
3. As usual, clustering method makes the centroid unstable while a new user joins. Although it is a effective method to locate user's group, it makes the members of user-group will unstable in any time.
4. When the RS perform user clustering, a user should belong to a specific group at least. Whereas a user cannot understand or realize the meaning of group which he/she stays.

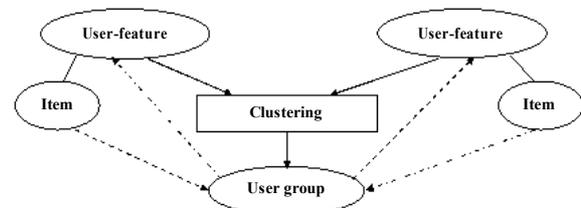


Fig. 2 Concept of collaborative filter analysis

2.3. Club member concept

In a real world, people may possess several interests or hobbies with different preference degree at the same time. This concept differs from traditional content-base approach and collaborative filter analysis. It emphasizes that a user can be recommended to join some interesting clubs or groups under rational conditions. According to the multi-club concept, a user can acquire information not only from the other members in the same club, but also from the other interesting clubs (as shown in Fig. 3). The club member concept can predict a user's interest intend and provide information meet he/she interests more flexible and effective in a RS. Several advantages of this method are described as follows:

1. Each club has clearly topic and meaning to represent itself, it is different from traditional ambiguity

in collaborative filtering analysis.

2. Once a user is classified to a specific club, he/she will stay in the club until he/she loss interest.
3. Each user can possess more interest choices with different preference degree. It can satisfy user's requirements for multi-interest.
4. A user can acquire information shared by the other members in the same club.
5. It is helpful for finding potential users from new coming users, and also eliciting new products consuming target from existed members.
6. Collaborative filtering and content-based are still support this concept, users can share information each other in the same club and acquire information that classified to this club by RS.

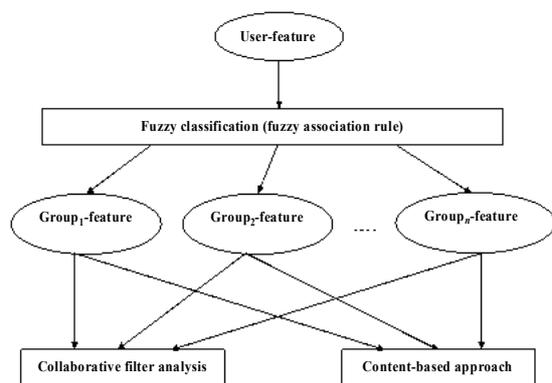


Fig. 3 Concept of club member

3. FUZZY ASSOCIATION RULE

In general, the interesting or preference of each user is difficult to describe clearly. In order to provide users multi-club service in the RS, fuzzy association rules are applied to classify users into appropriate clubs. For the sake of obtaining valuable fuzzy association rule for classification problem, data mining technology based on Apriori algorithm (Hu et al., 2004) is applied in this study.

3.1. Fuzzy partition methods

Fuzz sets and linguistic variables concepts were proposed by Zadeh (1965, 1975ab, 1976); they are feasible to represent the diversity of degree of the feature with different linguistic value in many applications (Hu et al., 2004). Formally, a linguistic variable is characterized by a quintuple (Pedrycz and Gomide, 1998; Chen and Jong, 1997; Zimmermann, 1991) denoted by $(x, T(x), U, G, M)$, in which x is the name of the variable; $T(x)$ denotes the set of names of linguistic values or terms, which are linguistic words or sentences in a natural language, of x ; U denotes a universe of discourse; G is syntactic rule for generating values of x ; and M is semantic rule for associating a linguistic value with a meaning. Using the simple fuzzy partition methods, each attributes can be partitioned by various linguistic values.

In this paper, each quantitative attribute is partitioned by simple fuzzy partition methods with triangle membership function. For example, the attribute "Age" is divided into 3 linguistic values (low, medium, high) and its membership function is predefined as follows (shown as Fig. 4). These membership functions can be represented by triangle fuzzy numbers. For example, the linguistic variable A_{med}^{Age} can be represented as $A_{med}^{Age} = (l, m, u)$, where m is the hedge value of linguistic variable A_{med}^{Age} .

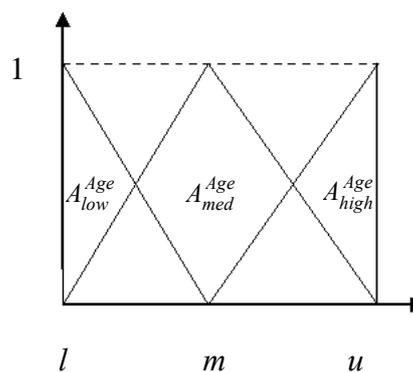


Fig. 4 Linguistic values of "Age"

$$m_{low}^{Age}(x) = \begin{cases} 1, & x < l \\ \frac{m-x}{m-l}, & l \leq x \leq m \\ 0, & x > m \end{cases} \quad (1)$$

$$m_{med}^{Age}(x) = \begin{cases} \frac{x-l}{m-l}, & l < x \leq m \\ \frac{u-x}{u-m}, & m < x < u \\ 0, & otherwise \end{cases} \quad (2)$$

$$m_{high}^{Age}(x) = \begin{cases} 0, & x < m \\ \frac{u-x}{u-m}, & m \leq x \leq u \\ 1, & x > u \end{cases} \quad (3)$$

In addition, each qualitative attribute is partitioned by a finite number of linguistic values in accordance with its categories. For example, the attribute "Gender" is divided into 2 linguistic values and its membership function is predefined as follows.

$$m_{male}^{Gender}(x) = \begin{cases} 1, & \text{if } x = \text{male} \\ 0, & otherwise \end{cases} \quad (4)$$

$$m_{female}^{Gender}(x) = \begin{cases} 1, & \text{if } x = \text{female} \\ 0, & otherwise \end{cases} \quad (5)$$

Each attribute is partitioned into appropriate linguistic value and obtained its membership in accordance with membership function, respectively. These linguistic

values are regarded as initial candidate fuzzy grids to generate the other large fuzzy grids and fuzzy association rule for classification in following tasks.

3.2. Fuzzy association rule

In order to generate valuable fuzzy association rule for classification, the large fuzzy grid concept is applied in this paper. If a database possesses k attributes and n tuples of member profile, the fuzzy classification rules are generated as follows.

Step 1. Perform the fuzzy partition to generate initial candidate fuzzy grids.

Step 2. Scan the database, and fuzzy support (FS) value is calculated as follows(Ishibuchi et al., 2001a; Ishibuchi et al., 2001b; Hu et al., 2002).

$$FS(A_{s_i}^{x_i}) = \frac{1}{n} \sum_{j=1}^n m_{s_i}^{x_i}(d_j^{x_i}) \quad (6)$$

where, x_i is the i th attribute, s_i is the term of degree in x_i , $A_{s_i}^{x_i}$ is the linguistic value of s_i in the i th attribute, $m_{s_i}^{x_i}$ is the membership function of s_i in the i th attribute, $d_j^{x_i}$ is the original value of the i th attribute in the j th tuple.

Step 3. If the FS value of linguistic value is larger than or equal to the user-specified min FS, it will be regarded as an eligible fuzzy grid.

Step 4. Those eligible fuzzy grids are combined to calculate their FS value as follows.

$$FS(A_{s_1}^{x_1}, A_{s_2}^{x_2}, \dots, A_{s_k}^{x_k}) = \frac{1}{n} \sum_{j=1}^n [m_{s_1}^{x_1}(d_j^{x_1}) \cdot m_{s_2}^{x_2}(d_j^{x_2}) \cdot \dots \cdot m_{s_k}^{x_k}(d_j^{x_k})] \quad (7)$$

Step 5. Go to step 3 to check the FS value of each fuzzy association rule. It will stop when only one rule is larger than or equal to the user-specified min FS. Otherwise, the processes between steps 3 to 5 will repeat again.

Step 6. Filter out ineffective fuzzy association rules with fuzzy confidence. The general fuzzy classification rule denoted by R is stated as

$$R : A_{S_1}^{x_1} \times A_{S_2}^{x_2} \times \dots \times A_{S_k}^{x_k} \Rightarrow A_{S_a}^{x_a} \quad \text{with } CF(R) \quad (8)$$

where, $x_a (1 \leq a \leq k)$ is the class label and $CF(R)$ is the certainty grade of R . It means that: if x_1 is $A_{S_1}^{x_1}$ and x_2 is $A_{S_2}^{x_2}$ and ... and x_k is $A_{S_k}^{x_k}$, then x_a is $A_{S_a}^{x_a}$ with certainty grade $CF(R)$. The fuzzy confidence (Ishibuchi et al., 2001a; Ishibuchi et al., 2001b; Hu et al., 2002) of R is defined as follows:

$$FC(R) = FS(A_{S_1}^{x_1} \times A_{S_2}^{x_2} \times \dots \times A_{S_k}^{x_k} \times A_{S_a}^{x_a}) / FS(A_{S_1}^{x_1} \times A_{S_2}^{x_2} \times \dots \times A_{S_k}^{x_k}) \quad (9)$$

When the $FC(R)$ of rule R is not larger than or equal to the user-specified min FC, the rule R will be regarded as ineffective fuzzy association rule and filtered out.

Step 7. Prune the redundant fuzzy association rules. If there exist two rules, denoted by R and S , possessing same consequence, and the antecedence of R is contained in S , then R is redundant and can be discarded.

4. EXAMPLE

There are 10 tuples of members profile in an exercise club as shown in Table 1. In this paper, the fuzzy classification rule is applied to classify a new coming user into multi-club with different preference degree. Hence, it is essential to locate effective association rule from the profile of existing member in each club. According to these classification rules, a new user can be classified into interesting clubs more effectively. The processes of generate fuzzy classification rule are described as follows.

Step 1. Perform fuzzy partition to transform original value of each attribute into linguistic value. We apply membership mapping table (shown as Table 2) to assign each education degree with a suitable membership in accordance with linguistic value, respectively. The other attributes are partitioned into linguistic values and obtained membership by membership functions. The linguistic value of each attribute is shown as Table 3. For example, the original value of attribute "Age" of member 10 is 43, and then membership degrees are 0.65, 0.35 and 0 for A_{high}^{Age} , A_{med}^{Age} and A_{low}^{Age} , respectively.

Step 2. Scan the database, and the FS of each linguistic value is calculated (shown as Table 4). For example, the FS of A_{high}^{Age} is calculated as.

$$FS(A_{high}^{Age}) = \frac{1}{10} (1 + 1 + 0.2 + 0 + 0.1 + 0.3 + 1 + 0 + 0 + 0.65) = 0.43$$

Step 3. Find the eligible fuzzy grids that FS value is larger than or equal to the user-specified min FS (In this case, the min FS is 0.3).

Step 4. In this case, the fuzzy classification rules aim to judge a new coming user whether interests this club or not. Therefore, the eligible linguistic value of "Preference" will be regarded as classification label to mining the association rules between preference degrees with other attributes.

Step 5. After the third scan, only one rule is larger than or equal to the user-specified min FS and the eligible large fuzzy grids are shown in Table 5.

Step 6. Calculate FC value of those eligible large fuzzy grids, and then filter out those ineffective fuzzy association rules that FC value is smaller than the

user-specified min FC(In this case, the min FC is 0.5).

Step 7. When the redundant fuzzy association rules are pruned, the fuzzy classification rule is remained only one as “If Age is medium, Education is high, then Preference is medium.”

Table 1 Member’s profile

Member	Age	Education	Income(K)	Preference
1	50	Doctor	56	0.96
2	56	Bachelor	67	0.89
3	34	Master	34	0.56
4	23	Bachelor	34	0.34
5	32	Master	23	0.52
6	36	Master	78	0.73
7	53	Bachelor	74	0.62
8	28	Master	32	0.55
9	25	Bachelor	55	0.47
10	43	Doctor	60	0.32

Table 2 Education membership mapping

Education	Top	High	Medium
Doctor	1	1	0
Master	0.8	1	0
Bachelor	0.5	0.8	0.3

Table 3 Linguistic value of each attribute

Attribute	High	Medium	Low
Age	(50,50,30)	(50,30,10)	(30,10,10)
Income	(90,90,50)	(90,50,20)	(50,20,20)
Preference	(0.8,0.8,0.4)	(0.8,0.4,0)	(0.4,0,0)

Table 4 FS value of each linguistic value

Attribute	High	Medium	Low
Age	0.43	0.51	0.07
Income	0.23	0.52	0.26
Preference	0.46	0.50	0.04

	Top	High	Medium
Education	0.72	0.92	0.12

Table 5 Fuzzy classification rules

Fuzzy association rule	FS(R)	FC(R)
$A_{top}^{Edu} \Rightarrow A_{high}^{pref}$	0.34	0.47
$A_{top}^{Edu} \Rightarrow A_{med}^{pref}$	0.35	0.49
$A_{high}^{Edu} \Rightarrow A_{high}^{pref}$	0.43	0.47
$A_{high}^{Edu} \Rightarrow A_{med}^{pref}$	0.46	0.50
$A_{med}^{Age} \Rightarrow A_{med}^{pref}$	0.33	0.64
$A_{med}^{Age} \times A_{high}^{Edu} \Rightarrow A_{med}^{pref}$	0.30	0.63

5. CONCLUSION

As information explosive grows, people must face the information overload more serious than before. RSs provide several solutions for people to acquire information meet their requirements. In order to achieve this goal, RSs need more effective and precisely method to improve the quality of information recommendation.

In this paper, we propose a club-member concept to allow a user can acquire more information from a variety of interesting clubs. In addition, fuzzy association rule based on data mining technology is applied to classify users into multi-club with different preference degrees in accordance with his/her personal features. Those fuzzy association rules can assist RSs to classify a new user into some suitable or interesting clubs for user. It is helpful for user to acquire more information from RSs more easily and effectively.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support of the National Science Council, Taiwan, China, under project numbers NSC 92-2213-E-212-008.

REFERENCES

[1] Chen, C.T. , W.S. Tai, “An information push-delivery system design for personal information service on the Internet,” *Information Processing and Management*, Vol. 39, Issue 6, pp. 873-888, 2003.

[2] Chen, S.M., W.T. Jong, “Fuzzy query translation for relational database systems,” *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 27, No. 4, pp. 714-721, 1997.

[3] Fang, X., O.R. Liu Sheng, “LinkSelector: A Web mining approach to hyperlink selection for Web portals,” *ACM Transactions on Internet Technology (TOIT)*, Vol. 4, Issue 2, pp. 209-237, 2004.

[4] Garcia, A.C.B., M. Ekstrom, H. Björnsson, “HYRIWYG: leveraging personalization to elicit honest recommendations,” *Proceedings of the 5th ACM conference on Electronic commerce*, New York, USA, pp. 232-233, 2004.

[5] Herlocker, J.L., J.A. Konstan, J. Riedl, “Explaining collaborative filtering recommendations,” *Proceedings on the ACM 2000 Conference on Computer Supported Cooperative Work*, Philadelphia, USA, pp. 241-250, 2000.

[6] Hu, Y.C., J.S. Hu, R.S. Chen, G.H. Tzeng, “Assessing weights of product attributes from fuzzy knowledge in a dynamic environment,” *European Journal of Operational Research*, Vol. 154, Issue 1, pp. 125-143, 2004.

[7] Hu, Y.C., R.S. Chen, G.H. Tzeng, “Mining fuzzy association rules for classification problems,” *Computers and Industrial Engineering*, Vol. 43, Issue 4, pp. 735-750, 2002.

- [8] Ishibuchi, H., T. Nakashima, T. Yamamoto. "Fuzzy association rules for handling continuous attributes," *Proceedings of IEEE International Symposium on Industrial Electronics*, Pusan, Korea, pp. 118-121, 2001b.
- [9] Ishibuchi, H., T. Yamamoto, T. Nakashima, "Fuzzy data mining: effect of fuzzy discretization," *Proceedings of the 1st IEEE International Conference on Data Mining*, San Jose, USA, pp. 241-248, 2001a.
- [10] Pedrycz, W., F. Gomide, *An Introduction to Fuzzy Sets: Analysis and Design*, Cambridge: MIT Press, 1998.
- [11] Ribak, A., M. Jacovi, V. Soroka, "“Ask before you search”: peer support and community building with reachout," *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, New Orleans, Louisiana, USA, pp. 126-135, 2002.
- [12] Shih, L.K., D.R. Karger, "Using urls and table layout for web classification tasks" *Proceedings of the 13th conference on World Wide Web*, New York, USA, pp. 193-202, 2004.
- [13] Sugiyama, K., K. Hatano, M. Yoshikawa, "Adaptive web search based on user profile constructed without any effort from users," *Proceedings of the 13th conference on World Wide Web*, New York, USA, pp. 675-684, 2004.
- [14] Wang, Y.F., Y.L. Chuang, M.H. Hsu, H.C. Keh, "A personalized recommender system for the cosmetic business," *Expert Systems with Applications*, Vol. 26, Issue 3, pp. 427-434, 2004.
- [15] Weng, S.S., M.J. Liu, "Feature-based recommendations for one-to-one marketing," *Expert Systems with Applications*, Vol. 26, Issue 4, pp. 493-508, 2004.
- [16] Zadeh, L.A. "The concept of a linguistic variable and its application to approximate reasoning," *Information Science*, Vol. 8, Issue 3, pp. 199-249, 1975a.
- [17] Zadeh, L.A. "The concept of a linguistic variable and its application to approximate reasoning," *Information Science*, Vol. 8, Issue 4, pp. 301-357, 1975b.
- [18] Zadeh, L.A. "The concept of a linguistic variable and its application to approximate reasoning," *Information Science*, Vol. 9, Issue 1, pp. 43-80, 1976.
- [19] Zadeh, L.A., "Fuzzy sets," *Inform. and Control*, Vol. 8, pp. 338-353, 1965.
- [20] Zhang, T., V.S. Iyengar, "Recommender systems using linear classifiers," *The Journal of Machine Learning Research*, Vol. 2, pp. 313-334, 2002.
- [21] Zimmermann, H.J. *Fuzzy Sets, Decision Making, and Expert Systems*, Boston: Kluwer Academic Publisher, 1991.

A Web-Based Decision Support Systems for the Chronic Deteriorating Diseases

Chichang Chang

Department of Information Management, YuanPei Institute of Science and Technology
HsinChu, Taiwan, China
threec@mail.yust.edu.tw

ABSTRACT

In general, most human physiological organs systems, which are constructed by collecting more than one part to perform either single or multiple functions. In addition, the successive times between failures are not necessarily identically distributed. More generally, they can become smaller (an indication of deterioration). However, if any organic deterioration is detected, then the decision of when to take the intervention, given the costs of diagnosis and therapeutics, is of fundamental importance. At the time of the decision, the degree of future human organic deterioration, which is likely to be uncertain, is of primary interest for the decision maker (for example, determining the prevalence of disease, doing a population survey, or measuring the level of a toxin). This paper develops a possible structural design of decision support systems by considering the sensitivity analysis as well as the optimal prior and posterior decisions. The proposed design of Bayesian decision support systems facilitates the effective use of the computing capability of computers and provides a systematic way to integrate the expert's opinions and the sampling information which will furnish decision makers with valuable support for quality decision-making.

Keywords: Aging Chronic Diseases; non-homogeneous Poisson process (NHPP); Bayesian Decision Theory; Decision Support Systems.

1. INTRODUCTION

Demographic shifts in the population will lead to a further increase in the proportion of elderly and consequently of people with chronic diseases. For example, almost 75 percent of the elderly (age 65 and over) have at least one chronic. About 50 percent have at least two chronic diseases [5]. In addition, aging is a strong socially appealing issue with many implications for users as well as providers of healthcare. In general, most human physiological organs, which are constructed by collecting more than one part to perform either single or multiple functions. However, the successive times between failures are not necessarily identically distributed. More generally, they can become smaller (an indication of deterioration). If any organic deterioration is detected, then the decision of when to take the intervention, given the costs of treatments and failures, is of fundamental importance. At the time of the decision, the degree of future organic deterioration, which is likely to be uncertain, is of primary interest for the decision maker. Naturally, gathering additional data will not always be economical. It is of special interest to determine analytically or numerically the conditions under which it will be worthwhile to collect additional information. Therefore, we propose a Bayesian decision process to provide a significantly improved methodology for dealing with the decision problems of physiological organs systems which can determine the conditions for taking the different actions, and thereby help the decision-maker maximize expected profit (or minimize expected loss).

2. MODELS FOR DESCRIBING AGING

In order to model aging in chronic diseases, the

non-homogeneous Poisson process (NHPP) was introduced since it seems more plausible for human physiological system consisting of many organs. The system failure process is time-dependent and its intensity function of the failure process is assumed to be of the form $\lambda(x) = \lambda_0 h(\beta; x)$, where λ_0 is the scale factor, β is the aging deteriorating rate, x is the elapsed time, and $h(\cdot)$ can be any function that reflects the deteriorating process. Suppose that the system has a planned lifetime (i.e., time horizon) T and let the decision has to be made at time t . The crucial two-action decision is whether at time t , the failure rate of the system will be too high (in which case some risk reduction action needs to be taken), or whether it will still be within an acceptable range (in which case we can keep operating the system according to the status quo). Another option is to gather additional information before the final decision is made.

Bayesian decision analysis has been well developed for decades [3,18], especially in the fields of statistical decisions [2,7], reliability engineering [4,9], quality management [1,16], and decision science [6,15]. The basic elements of the Bayesian decision process are as follows:

- (a) Parameter space $\Theta : \{(\lambda_0, \beta) | \lambda_0 > 0\}$, where λ_0 is the scale factor and β is the deterioration rate. Both parameters are uncertain and can be estimated through experts' opinions.
- (b) Action space $A : \{a_1, a_2\}$, where a_1 is the status quo, and a_2 is the risk reduction action. (We eventually expand this to consider a third possible action, the collection of additional information).
- (c) Loss function L : a real function defined on $\Theta \times A$. If we decide to keep the system operating, then the loss we face is $L(\theta, a_1)$; if we decide to take the risk reduction

action, then the loss we face is $L(\theta, a_2)$.

(d) Sample space S : The additional information available to be collected. For example, the successive failure times till the n . failure can be denoted as the likelihood function of the form

$$f_{x_1, x_2, \dots, x_n}(x_1, x_2, \dots, x_n) = \left[\prod_{i=1}^{n-1} \lambda(x_i) \right] \exp(-\Lambda(x_n)) \quad (1)$$

where $\Lambda(x) = \int_0^x \lambda(u) du$ is the mean number of failures by time x in the NHPP. The cost of collecting this additional information should also be reflected in the decision process.

The following terminology will be used throughout this paper:

C_A : the cost of a failure if it occurs.

C_R : the cost of the proposed risk reduction action.

C_I : the cost of collecting additional information.

ρ : the reduction in failure rate that would result from the proposed risk reduction action ($0 < \rho < 1$).

M : the expected number of failures during the time period $[t, T]$ under the status quo.

The decision variable we are dealing with is then the expected number of failures during the time period $[t, T]$, i.e.,

$$M \equiv M(t, T, \lambda_0, \beta) = \int_t^T \lambda(s) ds \quad (2)$$

Note that the expected number of failures M is itself a random variable, since it is a function of the two uncertain parameters λ_0 and β , and this is the case where Bayesian analysis can be effectively performed. Suppose that the risk reduction action will reduce the failure intensity by a fraction ρ , where $0 < \rho < 1$, then the expected number of failures in $[t; T]$, if the risk reduction action is taken is given by

$$\int_t^T \lambda(s)(1 - \rho) ds = (1 - \rho) M. \quad (3)$$

On the basis of the assumptions given above, we therefore have a two-action problem with a linear loss function, where the loss for taking action a_1 (i.e., continuing with the status quo) is $C_A M$ and the loss for taking action a_2 (i.e., undertaking the risk reduction action) is $C_A(1 - \rho)M + C_R$. The expected loss for the status quo is simply $C_A E\{M\}$, and the expected loss for the risk reduction action is $C_A(1 - \rho)E\{M\} + C_R$. A natural conjugate prior for the power-law failure model proposed in [10] is given by

$$f(\lambda_0, \beta) = K' \lambda_0^{m-1} \beta^{m-1} [\exp(-\alpha) y_m^m]^{\beta-1} \exp(-\lambda_0 c y_m^\beta) \quad (4)$$

and also a natural conjugate prior for the exponential failure model proposed is given by

$$f(\lambda_0, \beta) = K' \lambda_0^m \exp\left\{ \beta \alpha (m+1) y_m - \lambda_0 c \frac{\exp(\beta y_m) - 1}{\beta} \right\} \quad (5)$$

where K' and K'' are the normalizing constants, and the parameters c and α can be chosen to give the desired

values of expectations for λ_0 and β , respectively, and the parameters m and y_m can be chosen to give the desired degrees of dispersions for λ_0 and β , respectively [10,11,12,13]. Both the natural conjugate priors allow for dependence between λ_0 and β and have relatively simple closed-form expressions for their moments. Furthermore, the joint prior distribution about λ_0 and β in the Bayesian decision process can be straightforwardly derived.

3. THE DECISION SUPPORT SYSTEM

The Bayesian decision process mentioned in the previous section is capable of not only dealing with the uncertainties but also taking into account prior knowledge. However, this process cannot be easily performed; in particular, it requires the technology of numerical integration for carrying out complicated computations. To deal with the problem and to provide decision makers an efficient information system as well, we proposed a structural design of DSS to assist decision makers in making optimal decisions of minimum losses for deteriorating systems.

According to the interpretations about effective and efficient DSSs from [8,20], several required functions should be included in our study. First, the DSS needs to provide the optimal prior decision without using failure data for Bayesian updating. In such a case, the decision is based only on the uncertainties quantified by decision makers. Secondly, the DSS has to notify the decision maker whether collecting additional information is desirable or not. In some cases, the decision maker might not be confident or comfortable about the prior decision since the prior information is too vague. Therefore, collecting failure data could be another alternative before making the final decision. Furthermore, the DSS also needs to provide the optimal posterior decision. It should integrate the quantified prior information and the collected failure data. Finally, the DSS should have the ability to perform sensitivity analysis about each uncertain factor since decision makers might not be satisfied with the numerical values they applied in the system. The DSS has to allow decision makers change each uncertain entity (e.g. deteriorating rate), and therefore derives a range of such uncertain entities within which the optimal decision remains unchanged. Also, the ability of performing what-if analysis is crucial for the DSS [14,17,19]. It is of important interest for decision makers to see what the resulting decision will be once they change some parameters. According to the discussion above, the inputs to the DSS would be the uncertainties mentioned previously and the failure data (if available), and the outputs would be the optimal prior decision, the optimal posterior decision, and the results of the sensitivity analysis and what-if analysis. In order to perform the required complicated numerical integration for the decision process, the specifications of hardware and software should be closely considered. The hardware should have the ability to correctly and quickly respond to decision makers before they get impatient and

the software should be easily and reliably programmed and maintained.

The DSS includes three major processes, which are decision derivation, sensitivity analysis, and what-if analysis, respectively. These processes along with the information database are the essence of the DSS. Figs. 1 and 2 show the system follow chart and the system framework of the DSS, respectively [14]. The DSS has input, output, and process three major parts. The detailed information descriptions of each part are as follows:

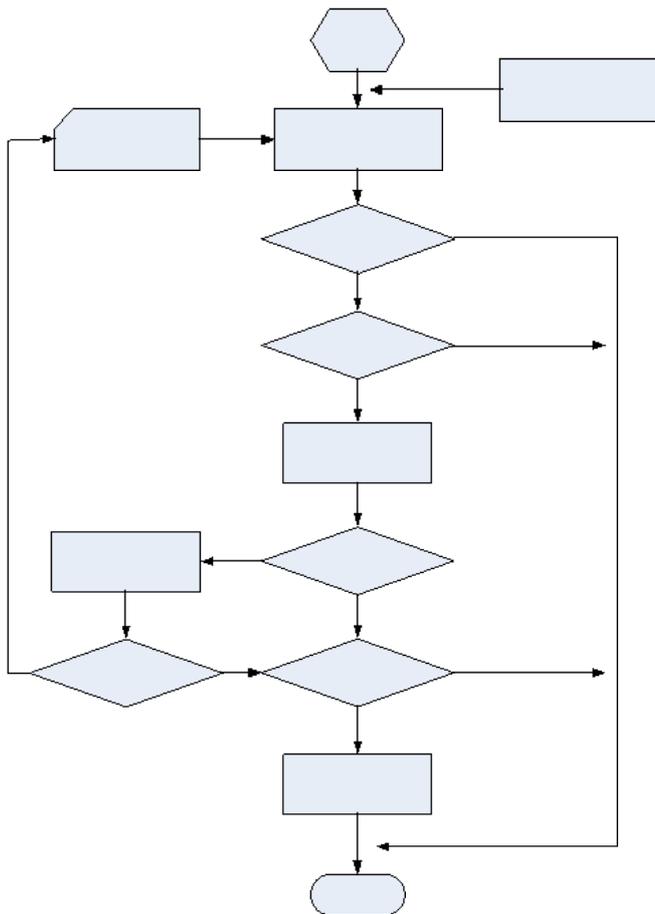


Fig.1. System follow chart of the DSS.

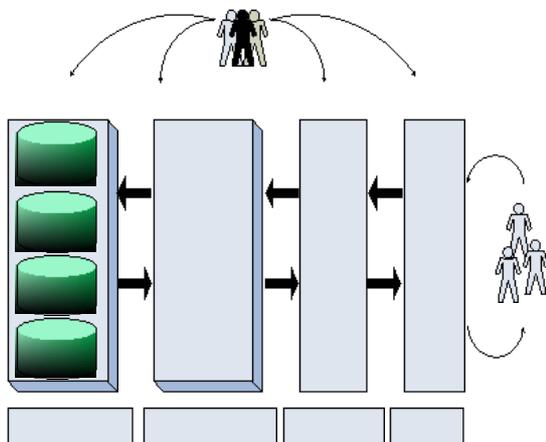


Fig.2. The system framework of the DSS.

Prior information part: The prior information has eleven elements and are described as follows:

- (1) *System lifetime*: The expected performing time of the physiological organs systems. (the measuring unit would be year.)
- (2) *Initial service date*: The date that the physiological organs systems first start.
- (3) *Decision time*: The actual time for decision makers to make the decision of whether maintaining the status quo or undertaking a risk reduction action.
- (4) *Cost of failure*: The cost or loss once the failure actually occurs.
- (5) *Cost of risk reduction action*: The cost for undertaking the risk reduction action.
- (6) *Risk reduction factor*: The fraction of the original function of the system that the risk reduction action can retrieve.
- (7) *Cost of collecting information*: The cost of collecting the failure data.
- (8) $E\{Scale\ Factor\}$: The expected value of the scale factor.
- (9) $SD\{Scale\ Factor\}$: The standard deviation of the scale factor.
- (10) $E\{Deterioration\ Rate\}$: The expected value of the deterioration rate.
- (11) $SD\{Deterioration\ Rate\}$: The standard deviation of the deterioration rate.

Sampling information part: The sampling information is for inputting the observed failure data.

Decision part: The decision part provides the optimal decisions that are suggested by the DSS. There are five output elements that can be valuable to decision makers for making the final decision. We introduce them as follows:

- (1) *Expected value of sampling information (EVSI)*: The EVSI can be treated as an indicator for determining whether to collect the failure data. In particular, if the EVSI were greater than the cost of collecting information applied in the prior information area, then collecting the failure data would be desirable; otherwise, collecting the failure data is not desirable.
- (2) *Prior $E\{\# of\ Failure\}$* : The expected number of failures for the remaining system lifetime under the status quo which is estimated by using the prior information only. This value shows the performance of the system if no risk reduction action is considered.
- (3) *Prior decision*: The suggested decision is based only on the prior information. It could be either maintaining the status quo or undertaking the risk reduction action. If collecting the failure data is evaluated as not desirable, then the prior decision suggested by the DSS should be considered as the optimal decision.
- (4) *Posterior $E\{\# of\ Failure\}$* : The expected number of failures for the remaining system lifetime under status quo which is estimated by using both the prior information and the failure data. This value shows the performance of the system if no risk reduction action is

Prior Information

Derive the Prior Decision
and Calculate the EVSI

considered when the prior knowledge of the system and the failure data are both applied to evaluate the system.

(5) *Posterior decision*: The suggested decision is based on both the prior information and the failure data. It could be either maintaining the status quo or undertaking the risk reduction action. Once the failure data is applied, the posterior decision suggested by the DSS should be considered as the optimal decision.

Once the decision area shows the decisions suggested by the DSS, the decision maker can perform further analysis to ensure the suggested optimal decisions are reliable. Sensitivity analysis can show the degree of importance for each prior parameter and study how they affect the optimal decisions. The DSS can provide one-way sensitivity analysis by using each element in the prior information as the changing factor. The results would be the ranges of the prior parameters that are of special interest in which the optimal decisions remain unchanged. The DSS also provides what-if analysis by changing the values of the prior parameters in the prior information and see if the optimal decisions are still unchanged or not.

4. CONCLUSIONS

This proposed structural design of DSS for risk management of deterioration in physiological organs systems can provide decision support techniques not only for taking action in the light of all available relevant information, but also for maximizing expected profit (or minimizing expected loss). It can deal with uncertain prior knowledge about the physiological organs systems by considering the optimal prior decision, the sensitivity analysis, and possibly, the optimal posterior decision (if actual failure data were available), and provide decision makers the effective support for quality decision making.

REFERENCES

- [1].Barlow RE, Irony TZ. The Bayesian approach to quality. In: Barlow RE, Clarotti CA, Squizzichino F, editors. Reliability and decision making. London: Chapman & Hall, 1993. p. 257–66.
- [2].Berger JO. Statistical decision theory and Bayesian analysis. New York: Springer, 1985.
- [3].Bernardo JM, Smith AFM. Bayesian theory. New York: Wiley, 1994.
- [4].Bunday BD, Al-Ayoubi ID. Likelihood and Bayesian estimation methods for Poisson process models in software reliability. International Journal of Quality and Reliability Management 1990;7(5):9–18.
- [5].Calkins E, Boulton C, Wagner E, et al. New ways to care for older people. Building systems based on evidence. New York: Springer; 1999.
- [6].Clemen RT, Reilly T. Making hard decisions with decision tools. PaciBc Grove, CA; Duxbury, 2001.
- [7].DeGroot MH. Optimal statistical decision. New York: McGraw-Hill, 1970.
- [8].Eom HB, Lee SM. Decision support systems research: a bibliography (1971–1988). European Journal of Operational Research 1990;46(3):333–42.
- [9].Erto P. New practical Bayes estimators for the 2-parameter Weibull distribution. IEEE Transactions on Reliability 1982;31(2):194–7.
- [10].Huang Y-S, Bier VM. A natural conjugate prior for the nonhomogeneous Poisson process with a power law intensity function. Communications in Statistics-Simulation and Computation 1998;27(2):525–51.
- [11].Huang Y-S. A decision process for deteriorating repairable systems. IIE Transaction 2001;33(6):479–85.
- [12].Huang, Y.-S. A Decision Model for Deteriorating Repairable Systems, IIE Transactions, Vol. 33, No. 6, pp.479-485, 2001.
- [13].Huang, Y.-S. and Bier, V.M. A Natural Conjugate Prior for The Non-Homogeneous Poisson Process With a Power Law Intensity Functions”, Communications in Statistic-Simulations and Computation. Vol.27, No.2, pp.525-551, 1998.
- [14].Mallach EG. Decision support and data warehouse systems. Boston: McGraw-Hill, 2000.
- [15].Rai-a H, Schlaifer R. Applied statistical decision theory. Cambridge, MA: Massachusetts Institute of Technology Press, 1968.
- [16].Singpurwalla ND. Design by decision theory: a unifying perspective on Taguchi’s approach to quality engineering. In: London: Chapman & Hall, 1993. p. 267–72.
- [17].Turban E, Mclean E, Wetherbe J. Information technology for management. New York: Wiley, 2002.
- [18].Winkler RL. Introduction to Bayesian inference and decision. New York: Holt, 1972.
- [19].Yamada S, Hishitani J, Osaks S. Software reliability measurement and assessment based on nonhomogeneous Poisson process models: a survey. Microelectronics and Reliability 1992;32(12):1763–73.
- [20].Zachary W. A cognitively based functional taxonomy of decision support techniques. Human-Computer Interaction 1986;2:25–63

An Ant-based Clustering Algorithm in Data Mining

Yong Tang, Yongkai Ma

Management School, University of Electronic and Science of China, Chengdu 610054, China
{TangYong, Mayongkai}@uestc.edu.cn

ABSTRACT

A new algorithm ABC (Ant based Clustering) inspired from behavior of the real ants is proposed for clustering in data mining. ABC employs some ants clustering in the searching space. Different from some clustering algorithms, instead of picking or dropping data objects, artificial ants in ABC find and merge the similar data points in their vision range, also a following and randomly route choosing strategy is introduced. Computational simulation results show that ABC is efficient, and the clustering results can be achieved in short time and maintain stably.

Keywords: data mining, clustering, ant colony algorithm

1. INTRODUCTION

Clustering is an important area in data mining. The purpose of clustering is to cluster a great deal of objects into groups according to the rule that is "things of one kind come together". The similarity of objects in a cluster should be as big as possible while the difference of objects in different clusters should be as big as possible too. Clustering has been applied in many domains, such as spatial data analysis, image processing, marketing and pattern recognition etc. Wildly used clustering algorithms are K-means, density based DBSCAN and network based STING etc [1,2]. Our ant based clustering algorithm, which is inspired by the collective behavior of real ants can avoid some shortages like (1) There is no need to determine the number of clusters artificially in advance in ABC. However, in general algorithms, users have to set the number as an input parameter. This requires rigorous domain knowledge for users, and on the other hand, the final clustering result is every sensitive to the initial number of clusters set in advance. But the truth is that, the number of clusters is not obvious especially when the data point is in a high-dimension space or the cluster rule is hard to tell. While in ABC, the clustering is going automatically, namely ABC might find the number of clusters by itself and when the clustering standard is chosen properly, the number of clusters and the results are credible and stable. (2) ABC can find clusters in any shape. Clustering algorithms based on the Euclidean distance output sphere-shape clusters, which is not a good way to describe the spatial distribution of data points in the space. In ABC, we introduce a clustering space or can be called as searching space. The position of a data point in the clustering space has no relations to its position in the high-dimensions data space. We randomly project data points in the data space into the clustering space, after doing that, we cluster them and the final distribution is our output, which indirectly indicates the distribution of data points in the original data space. (3) ABC can achieve the global optimization. Some algorithms are easily to stick into local optimizations. The ant based clustering algorithm

proposed in our research is a typical nature-inspired bionic algorithm. Artificial ants in the algorithm simulate the clustering behaviors of real ants to do random searches in clustering spaces, and taken the data points as the objects needed to be clustered. Every ant chooses a similar object in its vision range to move, meanwhile update the amount of the pheromone on the route to guide other ants. Repeatedly, the similar data points are clustered, when the count of iterations is reach an initial number or the error is less than an initial value we output the result. This algorithm does not require domain knowledge, and is able to generate the global clusters automatically. (4) ABC can process large amount of data in batches. General algorithms need to process all the data in one time, which is unrealistic when the data scale is large for the space and time limitations. Because ABC is a self-heuristic algorithm, it can process new inputs based on present clustering results, and adjust the clustering results dynamically.

2. A DESCRIPTION OF CLUSTERING PROBLEMS IN DATA MINING

First, we give a brief describe of the clustering problems. Data collected from different domains are usually in high-dimensions. And the distribution is what clustering concerns. Mathematically, we have n discrete data points X_1, \dots, X_n , which scattering in the d dimensions space \mathfrak{R} and for every point $X_i = (X_{i,1}, \dots, X_{i,d})^T$. Then we can define the distance between any two data points X_i and X_j as

$$d_{X_i, X_j} = \left(\sum_{k=1}^d p_k (X_{ik} - X_{jk})^2 \right)^{\frac{1}{2}}, \text{ where } p_k \text{ is the}$$

weight for the k dimension of data point X_i indicating the contribution of the k dimension under a certain of evaluating standard. To measure the result of the clustering also to see if the satisfaction of the clustering is met or not, error ε is introduced. For example, in the K-Means algorithm, after getting K

clusters, we can calculate the geometrical centers of these K clusters as $C_i, i=1, \dots, K$ and their

corresponding errors $\epsilon = \sum_{j=1}^K \sum_{i=1}^{N_j} d_{x_i, C_j}$, where N_j

is the number of data points fall into the same cluster whose center is C_j . When the error ϵ is smaller than a set value, we accept the clustering results otherwise we adjust the clusters. Existing clustering algorithms includes Hierarchical methods, Partitioning Methods, Grid-Based Methods and Density-Based Methods. And there are still some problems should be considered: (1) the outliers (2) applying users' domain knowledge in the clustering analysis (3) clustering large amount of data.

3. AN ANALYSIS OF CLUSTERING ALGORITHM

Ant in natural world is kind of typical social insect. By observing the collection behaviors of some objects such as food and corpses of ant colonies, biologists find that even without a centralized control mechanism, a colony of ants can do collections together in a very efficient way. This inspires us to deal our data points by simulating the collective behaviors of ant colony. Deneubourg [3] proposed a model to describe the clustering behaviors of ants, the basic idea is that ant can pick up and drop down objects with a probability rule as $P_{pickup} = (k_1 / (k_1 + f))^2$ and $P_{drop} = (f / (k_2 + f))^2$ for picking up and dropping down respectively. And f is the coefficient describes an ant's sensitive ability of its environment, k_1 and k_2 are limitations. Lumer and Faieta [4] improved this

model by using $d_{O_i, O_j} = (\sum_{k=1}^d (O_{ik} - O_{jk})^2)^{\frac{1}{2}}$ to

denote the Euclidean distance between two objects O_i and O_j . And the environmental parameter of object

O_i in a range of s is defined as $f(O_i) = \frac{1}{s^2} \sum_{O_j \in Neighbour_{s^2}(i)} [1 - \frac{d(O_i, O_j)}{\alpha}]$ We see

that if object O_i is similar to its environment, $f(O_i)$ will be relatively big, otherwise $f(O_i)$ will be small.

Thus the probability for picking up and dropping down is redefined as: $P_{pickup}(O_i) = (k_1 / (k_1 + f(O_i)))^2$ and

$P_{drop}(O_i) = \begin{cases} 2f(O_i) & \text{if } f(O_i) < k_2 \\ 1 & \text{if } f(O_i) \geq k_2 \end{cases}$. again k_1

and k_2 are limitations. Kuntz [5,6,7] applied this method in VLSI and Wu [8] did related work in the

clustering of WEB document. In the model of Deneubourg [3], Lumer and Faieta [4], at first, data points are randomly projected onto a 2-dimension plane, notice that the positions of data points on the plane have nothing to do with their positions in the data space. Then artificial ants do random movements, they pick up a data objects with probability P_{pickup} and drop data objects with P_{drop} . After some runs, the algorithm output the clustering results. The potential problem this kind of algorithm may have is that it might get stick in a halfway and fail to get data points clustered thoroughly, in other words, similar data points which should be clustered in a group, form several sub-clusters in the space and cannot be clustered together in a cluster any more. The reason for this is that $f(O_i)$ is used to denote the similarity of O_i to its local environment, when O_i is surrendered by some similar data points, the probability for O_i to be picked up will be small, consequently O_i might not be picked up any more, which means the O_i is fixed on the 2-dimension plane, its position will not be changed. Essentially, this result is locally optimized rather than global optimization we are seeking for. Monmarche [9] combined this kind of algorithm with K-means to achieve a fast clustering speed. Ant based clustering uses lots of ants searching in the space randomly with an enforcement of the pheromone on their routes to guide other ants. This positive feedback able the algorithm to guild further searches based on the already available searching experience and knowledge. The original model is first proposed by Dorigo [10] and performs very well in solving TSP and JSP, etc [11]. Yang [12] defines the mount of the pheromone between objects O_i and O_j

as $\tau_{O_i, O_j}(t) = \begin{cases} 1, & d(O_i, O_j) \leq r \\ 0, & d(O_i, O_j) > r \end{cases}$, and O_i is

clustered into O_j with a probability

$P_{ij} = \tau_{ij}^{\alpha} \eta_{ij}^{\beta} / \sum_{s \in Neighbour(i)} \tau_{sj}^{\alpha} \eta_{sj}^{\beta}$, where τ_{ij} denotes

the mount of pheromone, η_{ij} is the knowledge of the environment, which can be defined as

$\eta_{ij} = 1 / d(O_i, O_j)$. Obviously, the smaller the

distance between O_i and O_j , the more the mount of the pheromone, namely the higher the probability O_i and O_j been clustered in a same cluster.

4. NEW ANT BASED CLUSTERING ALGORITHM, THE IDEA AND STEPS

In our research we propose a new ant based clustering

algorithm. With a closer look at the behavior of real ants we believe that the vision ability of ants is limited in a certain range, namely their eyesight is limited, we use r to present the range ants can see. And ants are different, so they should not be treated in a same way. In our algorithm, some ants are big and some are small, namely their range of eyesight is in a direct proportion to the number of data points the $Cell_i$ has where the

ant occupies, then we get $r = k \sqrt{\frac{l^2}{N}} / 2, k = 1, 2, \dots,$

k for the number of data points been clustered on $Cell_i$, l is the width of the plane, N is the number of ants. So, an ant will see farther if $Cell_i$ has more data points. We let bigger ants search first in our algorithm, which is described as following:

(1) Randomly projects data points onto a 2D plane. One cell can only have one data point at most.

(2) Randomly place m ants, every ant carries the data points on a $Cell_i$.

(2.1) For ant Ant_i , we calculate its eyesight range

$r = k \sqrt{\frac{l^2}{N}} / 2, k = 1, 2, \dots,$ where k is the number

of data points been clustered on $Cell_i$.

(2.2) Calculate the average center of the data points on $Cell_i$ where ant Ant_i is. The center is

$\bar{C}_i = \frac{1}{k} \sum_{j=1}^k O_j$. If ant Ant_i finds there are data

points on cell $Cell_j$, and $d(\bar{C}_i, \bar{C}_j) \leq d$, where d

is a gate-value, then $Cell_j$ is added into a candidates

set *Candidate*. After the search, ant Ant_i finds out the most similar cell $Cell^*$ then merges all the data

points carried by Ant_i with the data points on $Cell^*$ and at the same time, we update the amount of

pheromone on the path between $Cell_i$ and $Cell^*$ to

1, namely $\tau_{Cell_i, Cell^*} = 1$ which indicates that this path

has been walked by Ant . Ant Ant_i is always trying to follow the path, which has been taken by those ants

set off from $Cell_i$, in other words, it attempts to find

$Cell_k$, where $\tau_{Cell_i, Cell_k} = 1$. If this attempt is failed,

we randomly move the data points of cell $Cell_i$ to a

cell $Cell_r$ on the plane, to avoid mistakes we must make sure that there are no data points on $Cell_r$.

(3) When all ants finish their moves, we calculate the

total error $\epsilon = \sum_{j=1}^K \sum_{i=1}^{N_j} d_{x_i, c_j}$, once final condition is

met, namely $\epsilon \leq \epsilon_0$, we output the clustering results and stop, otherwise go back to 2 and continue. Of course, we can set a count number *Count*, when the iterations gets *Count* or there is no changes in the clustering results, we stop and output.

In (1), as other algorithms do, we randomly project data points onto the 2D plane, we must remember that the 2D plane is a search space, the positions of n-D data points in the data space have nothing to do with the positions in the 2D plane. In this algorithm, artificial ants can handle both single data points and data points sets as well. Thus the problem we mentioned before can be avoided, that is a single data point cannot be moved for the similarity to its environment. By introducing the pheromone, we make the system have a kind of memory so that former searching experience can be reused to guide following searches. Artificial ants have vision range and in the searching process adapting a route-following and random choosing strategy make them closer to real ants in the nature.

5. SIMULATIONS

We apply the discussed ant based clustering algorithm in simulations. In this research, we use IRIS [13] in UCI machine learning database as our test dataset. There are 150 data points in the dataset and every data point has four parameters. There are 3 clusters: Iris Setosa, Iris Versicolor and Iris Virginica, each cluster has 50 data points. Our simulation environment is: OS: Windows XP, CPU: P4 1.5GHz, Memory: 128MB, Language is C. Our settings are: 10 ants, width for the plane is 15, *Count* is 100, *Count* will descend by one. After multiple simulations, we find that the ABC is good: average running time is less than 10 seconds, the number of clusters found by ABC is 3, mistaken points is between 4 and 5. One of the best run, only one data point is mistakenly clustered that is the NO.124 data point should be clustered in Iris Virginica but is clustered in Iris Versicolor instead. Wu did similar work [8].

Fig.1- Fig.4 is the process in a run. Fig.5 is the process of a run. We see that when *Count* = 25, we get 3 clusters and the result maintain stably.

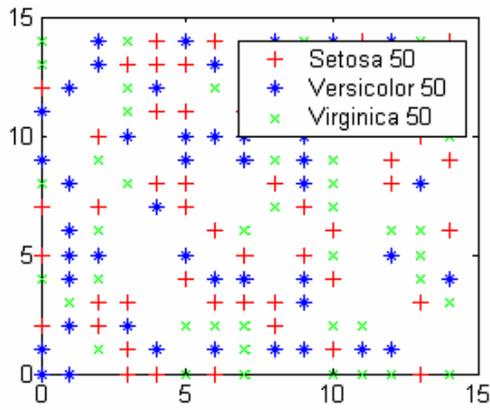


Fig.1 Count=0

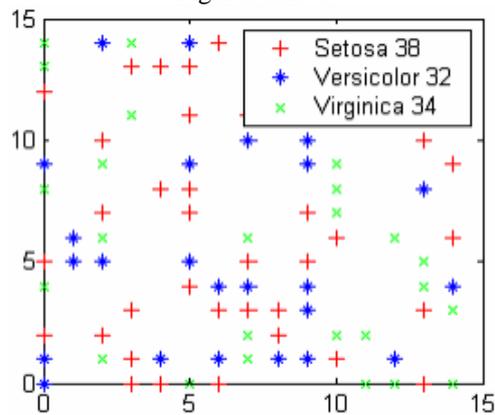


Fig.2 Count=5

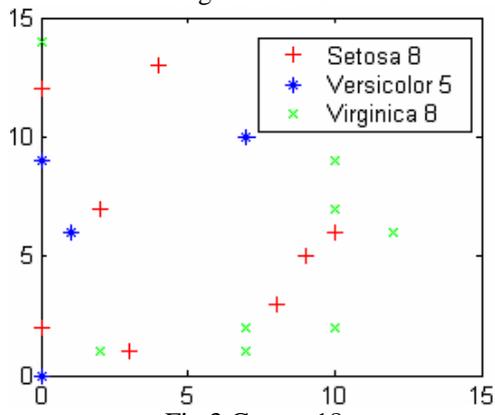


Fig.3 Count=18

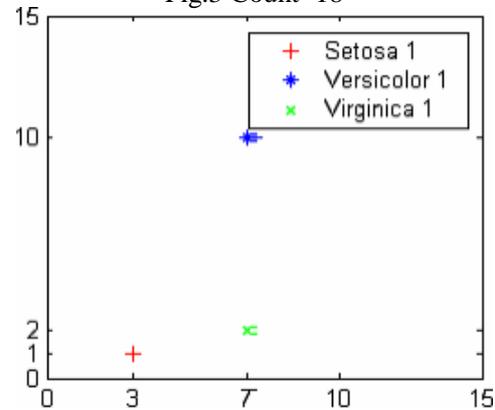


Fig.4 Count=25

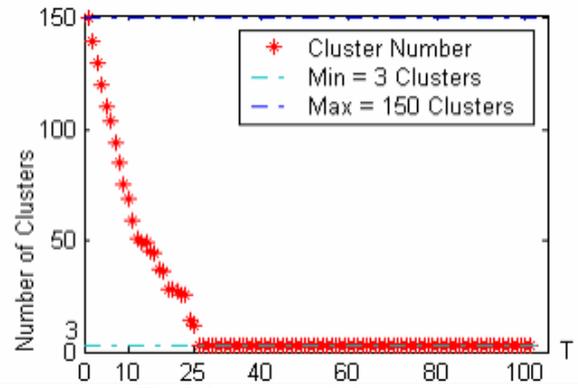


Fig.5 A process of clustering

6. CONCLUSION

Clustering analysis is one of the major areas in data mining and it has been widely applied in many domains. Clustering problems are usually in high dimensions and large scale. Nature-inspired algorithms have been used in many domains. We proposed an ant-based algorithm applying artificial ants in the searching space. Vision range concept is introduced with following strategy and random choosing strategy. The simulations show that the algorithm is fast and able to find the number of clusters automatically. As further work, we can consider: (1) At first we use ABC to find the number of clusters automatically, then we use other algorithms to do further clustering. (2) In ABC, we introduce a less similar data points kicking strategy to assure the correctness and reliability.

Applying nature inspired algorithms like ant colony algorithm to complex problems such as multiple objectives optimization, general function optimization and clustering analysis is something of swarm intelligence. As we have mentioned before, the original ant colony algorithm can solve every complex combinatorial optimization problems especially those can be abstracted into network optimization problems such as TSP etc. Swarm intelligence has a very important feature that is despite of the lacking of a centralized controlling mechanism or information spreading ways in certain structural forms ant colony can do things very effectively. Economic and social systems on complex networks can learn a lot from the collective behaviors of ant colony. The route choosing strategy used in the multiple objectives optimization inspires us to consider the decentralized collective behaviors in a brand new way. In other words, the artificial ant colony is a typical multiple agents system and the ant based clustering algorithm will be a good choice in the network clustering.

REFERENCES

[1] Han J., Kamber M. *Data Mining: Concepts and Techniques*, Beijing: Higher Education Press, 2001.
 [2] Jain A. K., Murty M. N., Flynn P. J., "Data Clustering: A Review". *ACM Computing Surveys*,

- Vol. 31, No. 3, pp264-323, 1999.
- [3] Deneubourg J. L., Goss S., Franks N. et al., "The dynamics of collective sorting: Robot-like ants and ant-like robots", *proceeding of the First International Conference on Simulation of Adaptive Behavior: From Animals to Animats*. MIT Press, Combridge, MA, pp356-363, 1991.
- [4] Lumer E. D., Faieta B., "Diversity and Adaptation in Populations of Clustering Ants" *proceedings of the third International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, MIT Press, Combridge, MA, pp501-508, 1994.
- [5] Kuntz P., Snyers D., "Emergent Colonization and graph partitioning" *proceedings of the third International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, MIT Press, Combridge, MA, pp494-500, 1994.
- [6] Kuntz P., Layzell P., Snyers D., "A colony of Ant-like agents for partitioning in VLSI technology", *proceedings of the fourth European Conference on Artificial Life*, MIT Press, Combridge, MA, pp417-424, 1997.
- [7] Kuntz P., Snyers D., Layzell P., "Astochastic heuristic for visualizaing graph clusters in a bi-dimernsional space prior to partitioning", *Journal of Heuristics*, No. 5, pp327-351, 1999.
- [8] Bin W., Zhongzhi S., "A clustering algorithm base on swarm intelligence", *proceedings of IEEE 2001 International Conferences on Info-tech & Info-net*, Beijing, China, pp58-66, 2001.
- [9] Monmarche N., "On data clustering with artificial ants", *Data Mining with Evolutionary Algorithms: Research Directions AAAI Workshop*, AAAI Press, pp23-26, 1999.
- [10] Dorigo M., Maniezzo V., Colomi A., "The ant system: optimization by a colony of cooperating agents" *IEEE Trans. System man Cybernet. B*, Vol. 26, No. 1, pp29-41, 1996.
- [11] Dorigo M., Caro G. D., Stutzle T., "Ant algorithms", *Future generation computer systems*, Vol. 16, No. 8, ppv-vii, 2000.
- [12] Xinbin Y., Jinggao S, Dao H., "A New Clustering Method Based on Ant Colony Algorithm" *proceedings of the 4th World Congress on Intellegent Control and Automation*, Shanghai, China, pp2222-2226, 2002.
- [13] Fisher R. A., "The Use of Multiple Measurements in Axonomic Problems" *Annals of Eugenics*, 7, pp179-188, 1936.

An Interactive Web Meeting System Using Multicast Technology

Ook Lee¹

¹College of Information and Communications, Hanyang University, Seoul, Korea(South)
ooklee@hanyang.ac.kr

ABSTRACT

Using a dedicated line for video conferencing is usually costly. A solution to solve this problem is to use a web-based video conferencing technology. However the existing webbased video conferencing technology needs too much bandwidth which makes two-way exchange of images too slow in the web environment. In this paper, a new web-based video conferencing system is introduced which has overcome this problem by using multicast method. The technique called tunneling which can preserve bandwidth enables this system to do multicast with minimum delay. Thus instead of using dedicated line-based video conferencing, web-based two-way multicasting technology is a better alternative.

Keywords: video conferencing, IP channeling, multicasting

1. INTRODUCTION

This article explains a web-based two-way multicasting technology. Web technology is used for many communication purposes. People post messages on the web board, and do chatting on a web environment. Due to the development of web cameras, nowadays many web surfers engage in multimedia chatting. In other words, people can conduct video conferencing through a web page. This method is definitely cheaper than non-web-based video conferencing which usually utilizes dedicated lines or satellites. This mode is called unicast which can not be used for multiple presentation of moving images in the same screen. In unicast mode one can only see one screen where a single moving image is projected. Of course it is possible to add multiple screens, however this requires separate lines, i.e., multiple number of unicast video conferencing systems. But it is also true that the unicast method provides high quality picture and no delays in projecting moving images whereas existing web-based multicasting method carries significant delays in showing moving images, which can result in irritation and inconvenience for participants. Furthermore two-way, or interactive conversation in the existing web-based multicasting systems is often slow and limited to just two parties; realtimesimultaneous multi-party conversation is usually not available. In this paper, the new web-based multicasting system that overcame the problem of "delay" in projecting moving images and allow real-time simultaneous multi-party video conferencing. This is possible through a technology called "tunneling" which reduces the delay time at the negligible rate. This web-based interactive multicasting system can replace expensive non-web-based unicast-style video conferencing technology as well as slow web-based existing video conferencing technology. With this system, one can not only do video conferencing but also broadcasting any video over the web page where the whole screen is divided into sections of showing participants' moving images. In other words, teaching a distance course with this system will cost much less than existing TV-based

method. Webbased learning has been around already for quite a while. But the problem of lack of interactivity and delays of showing moving images hindered further development in real-time multicasting-based distance education. This new technology will contribute a lot to distance education area by providing much cheaper and faster solution that can be interactive and multiparty-participation-possible. Web-based distance education is one of the most promising area in electronic commerce. And the success of web-based learning relies on the maturity of information technology infrastructure of the nation. There are studies showing the fact that the information technology infrastructure for electronic commerce is a key factor for the success of electronic commerce [1][2]. For example, high-speed Internet service available for most of the nation's population is a critical success factor for the rapid rise of electronic commerce in South Korea. Korea has been recognized worldwide as one of a few nations with highly developed information technology and Internet infrastructure. As for the highspeed Internet usage rate, Korea is reported as the number one country in the world. More than 10 million users subscribe to high-speed Internet services via such means as ADSL, VDSL, community LAN, and satellite. When taking the total number of Korean population into consideration, the high-speed Internet subscription rate in Korea is roughly 21% according to the Korean government ministry news [3]. This is a remarkably high usage rate when compared to those of other OECD countries since the average rate of highspeed Internet subscription in OECD countries is just 1.26% [4]. For many consumers in Korea, the Internet is just a part of their normal daily life. With the high-speed Internet available at very cheap price for almost all households, many electronic commerce firms flourish. Thus web-based multicasting video conferencing system is workable well in this highly-developed information technology infrastructure and web-based learning industry can be benefited much from this new technology.

2. WEB-BASED MULTICASTING VIDEO CONFERENCING SYSTEM

Web-based multicasting video conferencing system employs none Mbone router-supporting method. This method makes a different compared to the existing unicast-style video conferencing technology. In the existing technology, the system can not send multicasting data packet in M-bone supporting network environment. Regardless of support of Mbone by its routers, the web-based multicasting video conferencing system can send multicasting data packet by using tunneling technology. Figure 1 shows how tunneling technology works. In the existing method, data packets

are unicast from a server to a router and to a subrouter computer. In this way, two-way IP multicasting is not possible. However with IP tunneling technology the sub-router computer will recover data packets in webbased multicasting video conferencing system. This way also allows receiving data packets in reverse direction. Thus two-way IP multicasting is possible in the new system. What makes the new system superior is the fact that the time delay of showing moving images is less than 0.5 second, which means that the time delay is almost negligible. Figure 2 shows an actual example of video conferencing screen using the new system. In this example, 6 people are participating in two-way realtime video conferencing.

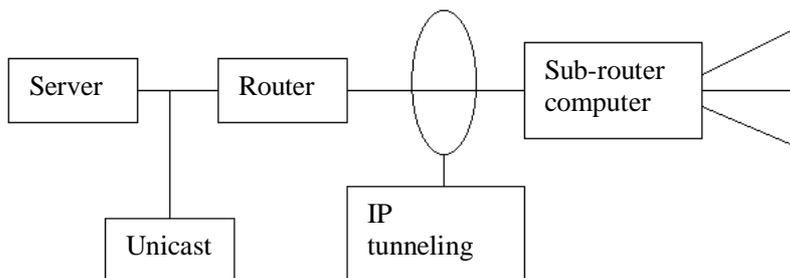


Figure 1



Figure 2

3. CONCLUSION

This system is a better system compared to the existing

system since it solves the bandwidth problem. In the existing unicast system multi-party meetings need too much network bandwidth. Namely the size of bandwidth

of the unicast system is bandwidth per person * # of people **2. For example, when 10 people are using the unicast system, the system requires $128 \text{ kbps} * 10^{**2} = 12,800 \text{ kbps}$, which is too much for the network to serve without significant time delay. On the contrary, in the web-based multicasting video conferencing system, the bandwidth is preserved using IP tunneling technology. Thus data packet transfer is effective through sharing. The bandwidth calculation for the new system is as following: bandwidth per person * 4 * (# of remote sites). For example, when 10 people are using the web-based multicasting video conferencing system, the system requires $128 \text{ kbps} * 4 * 1(\text{one location}) = 512 \text{ kbps}$. In one meeting room with two remote connection, $128 \text{ kbps} * 4 * 2(\text{two locations}) = 1,024 \text{ kbps}$. Thus no matter how many people are participating, the bandwidth of the new system is not affected by it. This is why the new system generates almost no time delay. The webbased multicast

video conferencing system can accomplish effective use of broadband Internet and provide high-quality video conferencing. This system will play an important role in promoting national IT infrastructure for distance learning industry which is rapidly growing in Korea.

REFERENCES

- [1] Weill, P. and Vitale, M. Place to Space: Migrating to eBusiness Models. Harvard Business School Press, 2001.
- [2] Weill, P. and Broadbent, M. Leveraging the New Infrastructure. Harvard Business School Press, 1998.
- [3] Ministry of Information and Communications, "Korea is the number one in high-speed Internet service", MIC news, www.mic.go.kr, 2004.
- [4] Samsung Economic Research Institute, Home page of SERI, www.seri.org, 2004.

An Investment Decision Support System (IDSS) for Identifying Positive, Neutral and Negative Investment Opportunity Ranges with Risk Control in Stock Markets

Wanli Chen

Faculty of Information Technology, University of Technology, Sydney, 2007, Australia
wchen@it.uts.edu.au

ABSTRACT

While there are a number of finance methods (fundamental analysis, technical analysis, contrarians' theory, risks management, etc) used in stock markets to help make investment decisions, they have different strengths and weakness. It is observed that these different finance methods are not being integrated by existing technologies in a systematic way, thus their performance for identifying investment opportunities is limited. In this research, I propose a systematic method (i.e. an IDSS system) to take advantage of and to optimize existing and newly proposed methods in order to obtain better investment performance, through identification and classification of Positive, Neutral and Negative investment opportunity ranges and related risks. This IDSS system will be mainly based on *Turning Point Model* and *Optimized AutoSplit* method, which help find hidden investment opportunities and risk variables, particularly a stock's unique trend. The key methodology is to use Decision Tree theory with finance knowledge. The IDSS system will be built on the top of *F-trade platform* which has been already developed by UTS Data Mining team and has a RDP structure, with agent-based distributed expert systems. Initial system evaluation shows that the system successfully identified investment opportunity ranges, outperforming the benchmark index and other systems.

Keywords: DSS, E-commerce, stock markets, RDP, turning point model, AutoSplit

1 BACKGROUND

While stock market is studied to be predictable, it has a number of finance methods with different strengths and weakness to identify investment opportunities. It is necessary to integrate them in a systematic way to take advantage of their strengths, but literature shows that few IT systems have accomplished it. Therefore, I design an Investment Decision Support System (IDSS) to integrate their strength in order to identify investment opportunities and risk factors.

Stock Investment Opportunity Identifiability: Research on the microstructure of security markets has concluded that the intersection of supply and demand curves for stocks determines stock price and accordingly liquidity [1]. For many years, stock investment opportunity identifiability has been studied by both academics and practitioners. It was observed that two competing explanations have been offered for it. Proponents of EMH (e.g. [2]) maintain that such identifiability results from time-varying equilibrium expected returns generated by rational pricing in an efficient market that compensates for the level of risk undertaken. Critics of EMH (e.g. [3]) argue that the identifiability of investment opportunities reflects the psychological factors, social movements, noise trading, and fashions or "fads" of irrational investors in a speculative market.

Necessity of Integrating Distinct Finance Methods: While there are so many finance methods (such as fundamental analysis, technical analysis, contrarians' theory, etc) in capital markets to help identify investment opportunities, they have different

characteristics and accordingly, strengths, weakness ([4], [5], [6], etc). For example, fundamental analysis focuses on stock's real value and its long term benefit, while technical analysis focuses on stock's price and volume changes and tries to achieve short-term profit. It is observed that these different methods are not integrated in a systematic way thus their performance for sound decision making are limited [7]. However, there is an increasing need of integration of different methods in stock markets, and it is becoming more and more common for practitioners to use different methods simultaneously to get an optimized result [8].

Current Information Technologies for Identifying Stock Investment Opportunities: There are a number of expert systems, knowledge engineering and other technologies to accomplish identification of stock investment opportunities. Many employ technical analysis momentum heuristics, and some use neural networks to learn patterns. Example include: neural networks learning from past price history [9]; neural network using knowledge of real world events [10]; fuzzy expert system combining several sources of information [11] rough sets to extract trading rules from price history [12]; data mining employing signal processing techniques [13]; grey theory and fuzzification [14]; and rules [15] or agents [16] using technical heuristics combined with other sorts of knowledge in an expert system framework. However, these information technologies mainly focus on technical analysis and few considered integrating different finance methods in a systematic way to achieve an optimized performance than any single method does; more importantly, they often ignore or

misunderstand importance of expert's domain knowledge and its distribution, therefore, their performance for sound investment decision is limited and they can not well meet the need of identifying investment opportunities for finance industry (i.e. brokerage firms).

Therefore, in this paper, I focus on designing an Investment Decision Support System (IDSS) for integrating both above-mentioned existing methods and proposed new methods in both finance and IT to help identify investment opportunities and risk factors, which can create a better performance than the benchmark index and other systems. In section 2, I will provide a

detailed description of Investment Decision System (IDSS) and related methodologies, including *Turning Point Model* and *Optimized AutoSplit Method*, and how to use them to identify investment opportunities and risk. In section 3, I will introduce IDSS implementation and its evaluation. Section 4 is conclusion.

2 INVESTMENT DECISION SUPPORT SYSTEM (IDSS) AND RELATED METHODOLOGIES

2.1 As the kernel of IDSS, a simplified architecture was depicted in Figure 1.

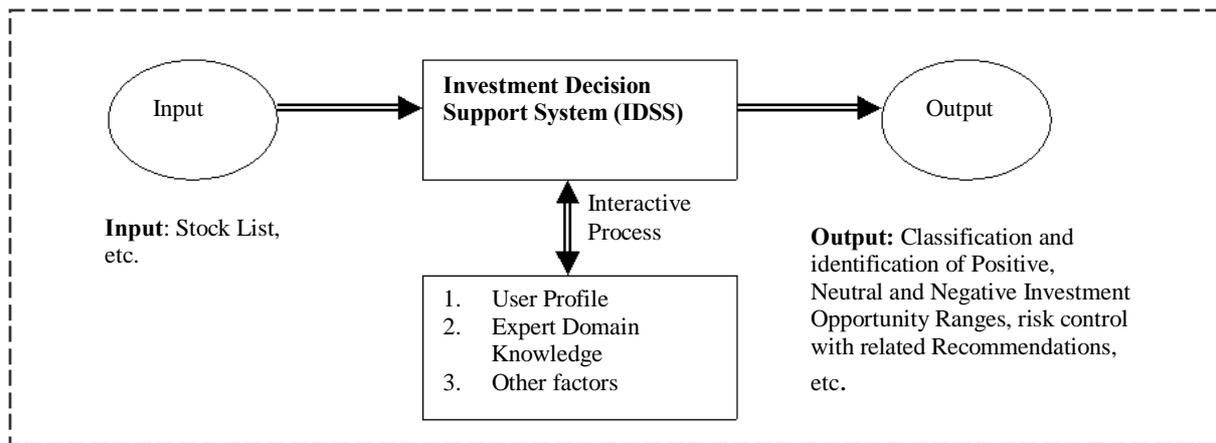


Figure 1: a simplified architecture of IDSS system

2.2 Turning Point Model

As shown in Figure 1, in order to build architecture of IDSS system, it is important to clarify three concepts: *Positive, Neutral and Negative Investment Opportunity*

Ranges. This Section provided a brief Introduction of *Turning Point Model* and the three concepts. A basic structure of "*Turning Point Model*" was depicted in Figure 2.

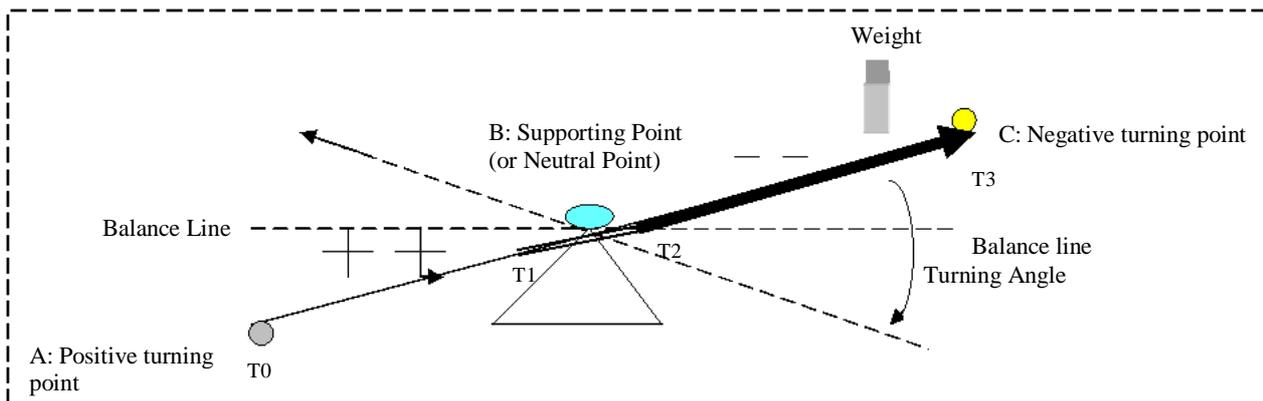


Figure 2: a basic structure of *Turning Point Model*.

From Figure 2, following methods can be derived to help identify investment opportunities and risks:

1) **Supporting Point (or Neutral Point):** This is the fuzzy point of indicating stocks' real value, where the market is rational and its effect is neutral (neither positive nor negative). Note that *Supporting Point* could move within a range when fundamental factors affect stock's real value, represented by Neutral Range ().

2) **Positive Turning Point:** This is the fuzzy point where market undervalues stocks to such an extreme extent that most investors run away and the rest are waiting on sell side of order book, but it starts turning positive because of contrarians' entry and others' subsequent following. The signals include: large (percentage) investors or/and previous large profit-makers start buying, combining with other related signals, such as their

historic behavior and trading patterns in similar situations, fundamental info (such as lowest P/E), how close they are to important fundamental info source, the confidence extent in which they can be followed, and medias' apparent exacerbation of stock's bad side.

- 3) **Accelerating Speed (++)**: Moving from Positive Turning Point, some leading investors notice contrarians' entry signal and start buying, then more and more investors follows. This trend moves on until it passes supporting point. *Accelerating speed* indicates the change extent of market movement and growth, which becomes faster and faster.
- 4) **Negative Turning Point**: when above trend moves beyond supporting point, most new investors are becoming over-optimistic about it and rush in. The lever becomes so unbalanced that a *negative turning point* comes. This is the fuzzy point where the stock value is extremely overvalued and most investors already rush in out of mood.
- 5) **Decreasing Speed (--)**: it indicates that the extent of change of market movement and growth when it is close to *negative turning point*, becomes slower and slower. The speed even decreases to become negative since most possible investors are already in, and investment temperature starts turning cooler from its hottest moment. Therefore, market

Decreasing Speed and cooling signals need be observed for investment, i.e., large (percentage) investors and/or previous large profit-makers procedurally leave the market and small investors now dominate a high percentage but their growth speed is decreasing even to be negative, and demand decreasing speed and supply increasing speed become equal which make no "momentum" any more.

- 6) **Relevant Balance**: it indicates the extent of out-of-balance of a point to balance line, and the model assumes that lever will finally bounce back to balance line.
- 7) **Turning Angle**: when lever is out of balance, it moves up and down and extent is *turning angle*.
- 8) **Weight (■)**: it indicates the weight of changes of a particular point, including multiplier of changes of volume, price, depth, individual (large) investors' market affectability etc.
- 9) **Weight Distance Multiplier**: it is a point's distance to *Supporting Point* multiplied by *weight*.
- 10) **Risk Factors Coefficiency**: the risk extent of a particular point can be identified based on its distance to *positive turning point*, *negative turning point* and *supporting point*, multiplied by weight, turning angle, speed, and other factors.

Figure 3 simplified Figure 2 and depicted how to use *Turning Point Model* in IDSS system.

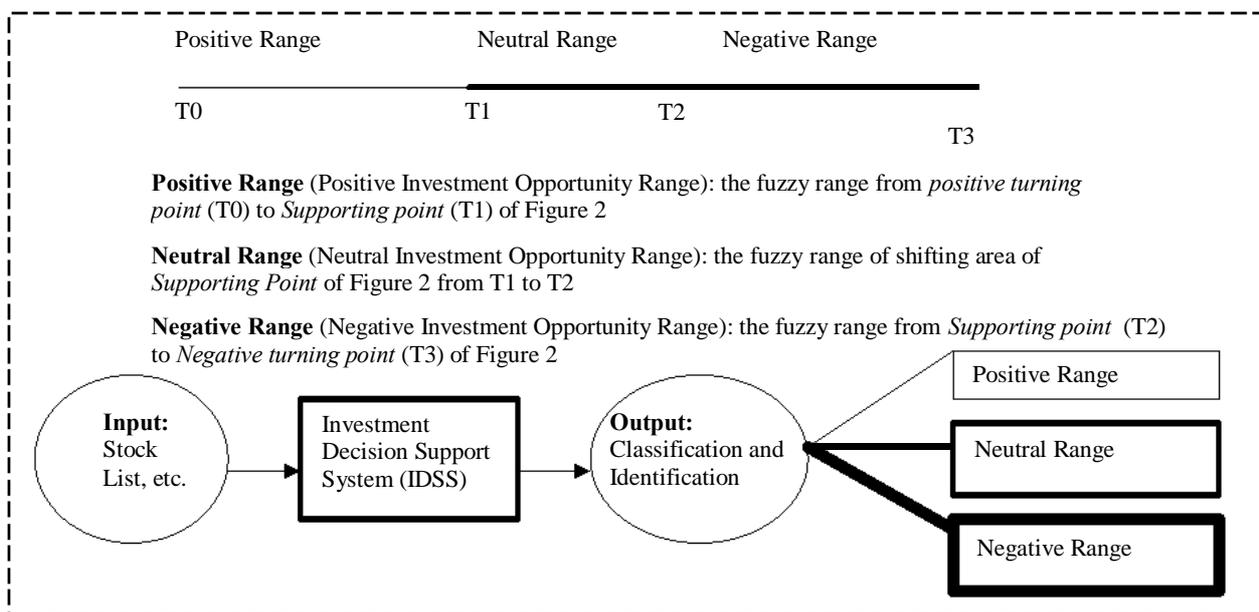


Figure 3: a simplified *Turning Point Model* and its usage in IDSS system

Benefits of using *Turning Point Model* and three concepts in IDSS system include:

- It avoids identifying an exact trading point which is difficult and unnecessary in real stock trading practice and it ignores frequent fluctuations in a short time; instead, IDSS system can use *Turning Point Model* to identify accurate different types of investment opportunity ranges for a mid-term time, which is much easier and suitable for stock trading practice and can create higher profit.
- The model integrates concepts and strengths of Fundamental Analysis, Technical Analysis, Contrarians' theory, etc. Based on it, IDSS can create better output than any single method does.
- The model also proposes new concepts and methods which can help better identify investment opportunities and related risks than most conventional methods do.

2.3 Optimized AutoSplit Methods

In order for IDSS to efficiently identify Positive, Neutral and Negative Investment ranges, I use *Optimized Autosplit Methods* to help identify these hidden investment features and risk variables. Pan et al [17] proposed AutoSplit Method and claimed that it is able to find features which are mutually independent and able to discover non-orthogonal features. However, it is mainly designed for image processing and not suitable for capital markets, thus it has much room for improvement. I proposed optimized AutoSplit methods to help better find investment opportunities and risk factors:

- 1) Setting Precondition of AutoSplit: AutoSplit method assumes that stock data distribution is not Gaussian (normal) and used a company as an attribute, without any proof. To improve it, I judge if stock data distribution is Gaussian as first step. If it is Gaussian, SVD is more suitable; if not, AutoSplit method is suitable. I use following methods to judge it:
 - Ø Using *Kolmogorov-Smirnov test to test Gaussian distribution*. I used GraphPad InStat software, when “ $P < 0.10$ ” (small P values), the population is unlikely to be Gaussian.
 - Ø Using *Shapiro-Wilk normality test or Dgostino-Pearson omnibus test*.
- 2) Integrating Multi-dimensional methods: in the experiment, AutoSplit only considers price thus data set is one dimensional. While in practice more types of data (i.e. company, price, volume, depth, spread or fundamental factors (P/E, Asset/liability), etc) need to be considered for prediction, the data set should be multi-dimensional. To improve it, I
 - Ø add multiple loops into algorithm to handle multi-dimensions, and
 - Ø Include (price, volume, any other data type) as a data cube.
- 2) Integrating with streaming methods: AutoSplit method does not consider change among time windows thus it is not quite useful for predicting trend. To improve it, I integrate following window models with it:
 - Ø *Landmark windows*: Aggregates are computed based on the values between a specific time point called the landmark and the present. Similarly, data stream can be divided into a “baseline window” and a “sliding window” and compare two data sets.
 - Ø *Sliding windows*: Aggregates are computed based on the last w values in the data stream. The size of a sliding window w is predefined.
 - Ø *Damped window*. The weights of data decrease exponentially into the past.
- 3) Reverse AutoSplit: While time is treated as major attributes and controls the order of data in AutoSplit method, reversely I treat time as a common attribute and use other attribute (such as price) to control the order and compare the results

to see new patterns (say, what time patterns shows at a particular price range).

- 4) Improving the experiments: Experiments can be improved in following ways:
 - Ø While it claims h_1 and h_2 represents the general trend and internet bubble of share price series, it can be proved by including DJIA index and Nastaq index (or IT industry particular index) as an attribute. If the claim is right, DJIA and Nastaq index should have highest
 - Ø To get richer or more interesting experiment results by doing following tasks:
 - a. Using intraday trading data as data source rather than weekly closing price;
 - b. Using real-time data as data set;
 - c. Conducting “auto-split” first within an industry, then across industries;
 - d. Findings of hidden variables should relate to domain knowledge;
- 5) Improving analysis of experiment results: Current AutoSplit focuses on major variables and ignore less significant ones, but less significant variables can have important impact on stock market and investment. I improve analysis of experimental results in following ways:
 - Ø To see if less significant hidden variables disclose more important information;
 - Ø To detect bursts of activities or abrupt changes in real time;
 - Ø To mine outlier or unusual patterns in stream data;
 - Ø To make the number of hidden variables more accurate: AutoSplit assumes that the number of hidden variables is the same as the number of attributes and it also use the *Whitening* step to control the number, thus it may ignore or mix important variables. Therefore, I assess the assumption first to make the number of hidden variables accurate.

Initial experimental results show that above methods are of research value and can play an important role in the success of IDSS, for instance:

- They efficiently identify a stock’s unique trends which only belong to the stock itself.
- They more efficiently help find hidden important investment opportunity variables and risk factors for sound investment decisions than conventional methods do.

2.4 a Sample of Investment Decision Support System (IDSS)

Based on the architecture of IDSS system (Figure 1), concepts of *Positive, Neutral and Negative Investment Opportunity Ranges* (Figure 2 and Figure 3) and *Optimized AutoSplit Model*, a sample of *Investment Decision Support System* (of identifying Positive Point and Range) was roughly depicted in Figure 4. Note that

this sample shows Decision Tree methodology broadly, but in details and implementation, IDSS uses fuzzy

rules rather than Boolean rules.

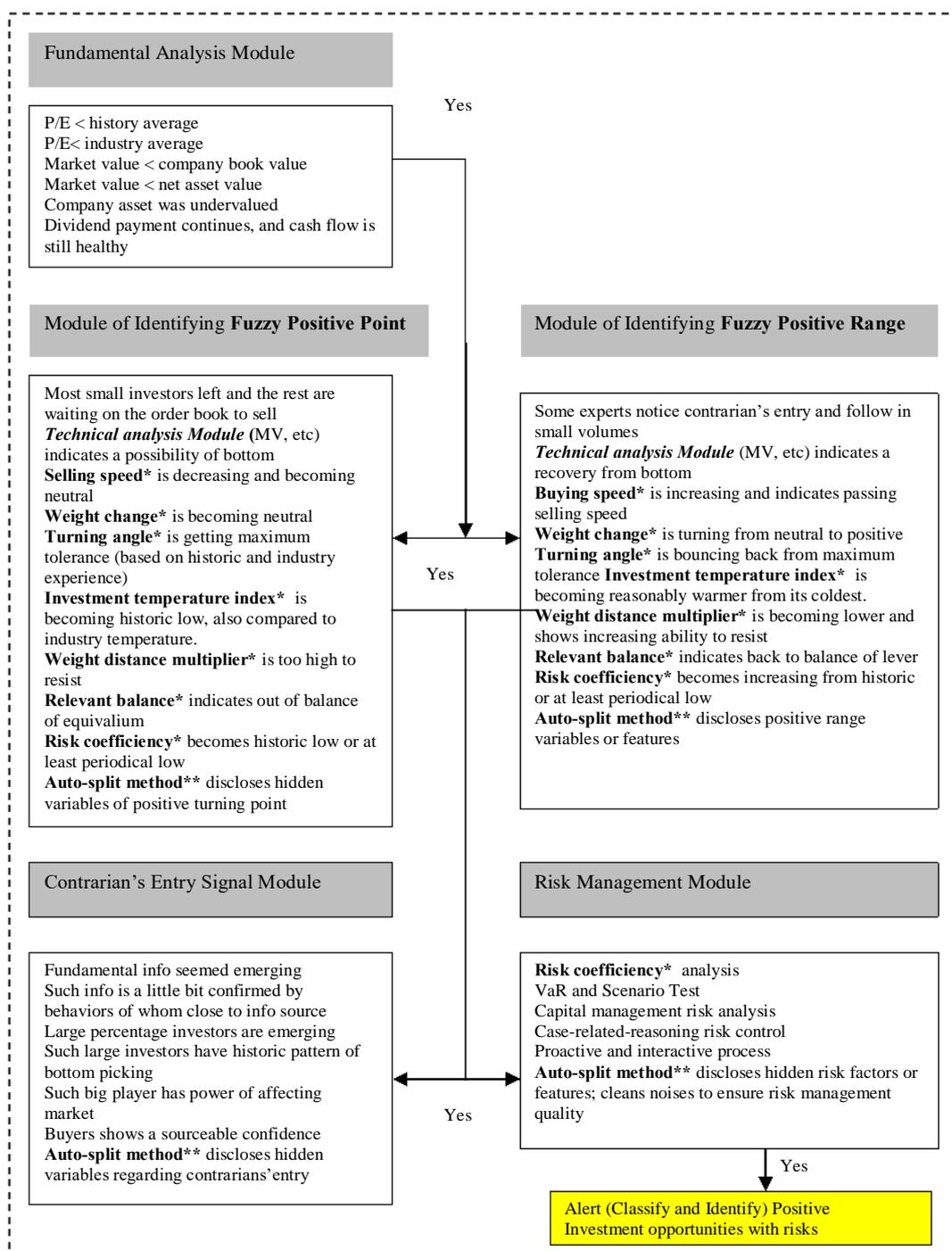


Figure 4: a sample Investment Decision Support System for identifying fuzzy Positive Investment Opportunity Ranges with Risk Control. * parts come from Turning Point model; ** parts come from AutoSplit method.

As shown in Figure 4, IDSS system is constituted by six modules, including Fundamental analysis module, Technical Analysis module, Identifying Fuzzy (Positive or Negative) Turning Point module, Identifying fuzzy (Positive or Negative) Investment Range module, Contrarian's Entry Signal module and Risk Control module. All these six modules are totally implemented

on existing Capital Market Cooperative Research Center (CMCRC) project F-trade platform, thus they will greatly benefit from strengths of F-trade platform. On the other hand, IDSS system extends the functions of F-trade platform and makes it more powerful. Their relationship was depicted in Figure 5.

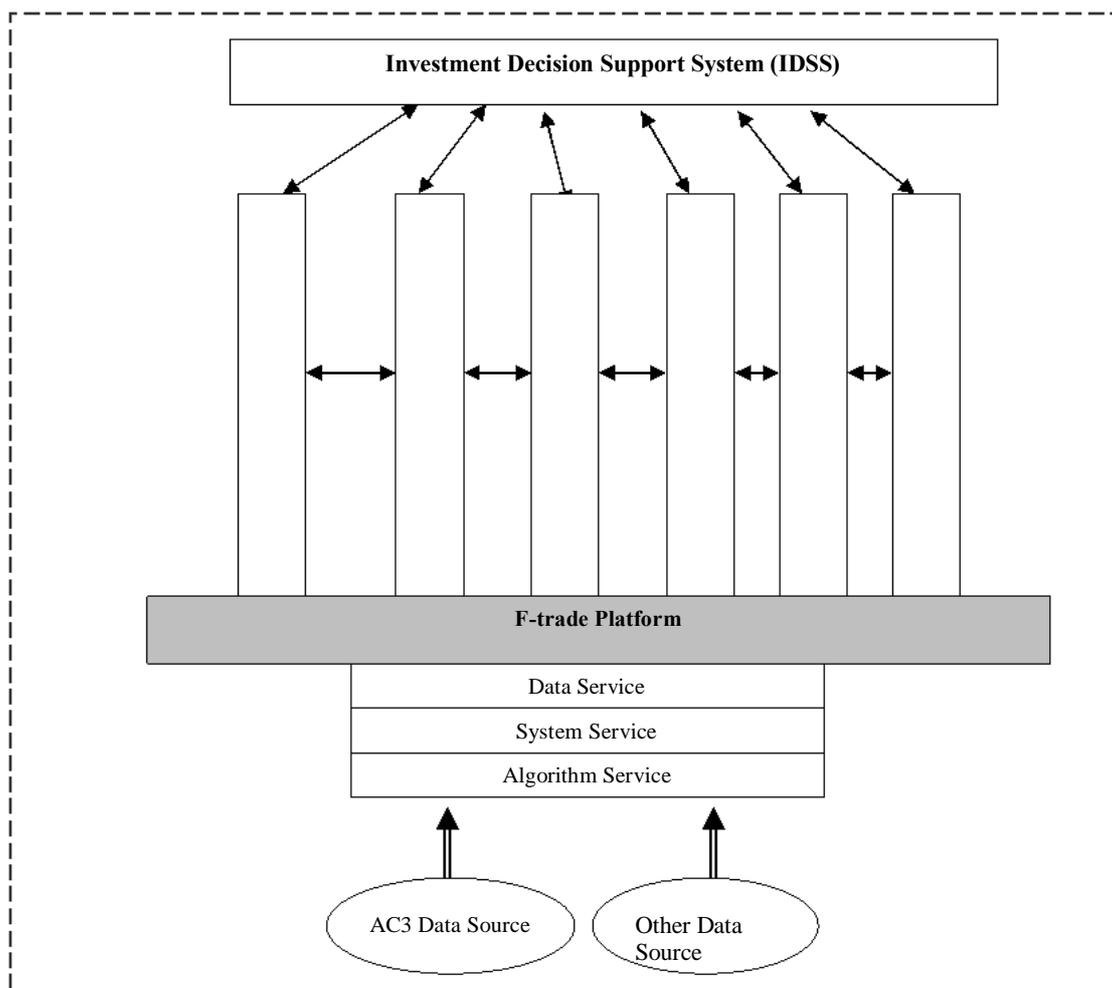


Figure 5: Relationship of IDSS System Modules and CRC project *F-trade platform*

Functionalities of six modules under IDSS system are described in this section:

- *Fundamental analysis module:* This module analyses fundamental factors which determines and affect a stock's real value. As show in Figure 4 & 5, the module may consider financial ratios such as P/E, financial figures such as book value, net asset value, cash flow, dividends payment and so on. It is mainly used to assess if the stock is undervalued or overvalued and the extent.
- *Technical analysis module:* This module uses existing technical analysis methods and algorithms such as Moving Average to evaluate the condition and trend of stock movement.
- *Identifying Fuzzy Turning Point module:* In this module, concepts and methods of *Turning Point Model* will be used to help assess if the stock is at fuzzy Positive Turning Point (in figure 4) or Negative Turning Point. It uses methods of Selling speed, Weight change, Turning angle, Investment temperature index, Weight distance multiplier, etc. Additionally, I adopt *Optimized Auto-split method* to help identify hidden variables and features of fuzzy Turning Point.
- *Identifying Fuzzy Investment Range module:* This module adopts same methods used by above module, but they are used to assess if stock is in fuzzy range, rather than turning points.
- *Contrarians' Entry Signal module:* Contrarians have the power of affecting market thus their entry means a turnaround of market. Not only quantitative methods (i.e. transaction details on orderbook, hidden variables of *Auto-split method*), but experts' experience and judgments work.
- *Risk Control module:* As risk and return are closed related, risk control is closely related to above modules. Not only traditional risk control methods (including VaR, Stress test, capital management, etc), but new methods from *Turning Point model* and *Auto-split method* work.

While IDSS system (top module) depends on above six modules, it has much more important roles. For example, IDSS will decide which of above six modules and other knowledge source will be used, how to use, and how to integrate and optimize the results from individual module to get an ideal output (i.e., classification and identification of *Positive, Neutral and Negative Ranges*). In follow section, I use a RDP model to depict a detailed system structure of IDSS.

Figure 6 used a RDP (Requestor-Dispatcher-Provider) model to depict the structure of IDSS. Simply speaking, once *Requestor* sends users' request to *Dispatcher*, *Dispatcher* will choose appropriate *Provider*, and then

Provider will look into model base, knowledge base, data base and expert domains to find appropriate solution. On a reverse way, the solution will be passed to users. It includes:

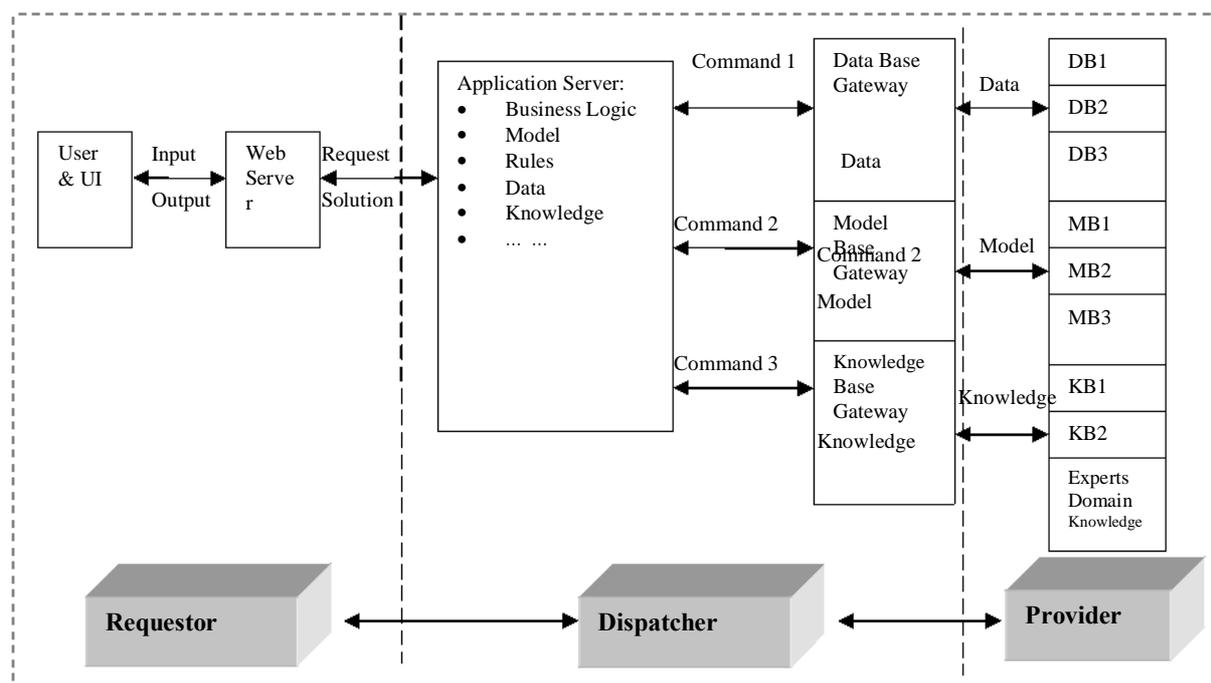


Figure 6: a System Structure of IDSS by Using RDP Model

(1) **Investment Requestor:** It contains user interface and web server and covers all aspects of the communications (i.e., input and output, problems and solutions) between a user (i.e., brokers, individual investors) and the IDSS; as a result, it derives much of the power, flexibility, and ease of use of IDSS. For example, it could send users' request of classifying a particular stock (i.e. BHP) at a particular time point (i.e. right now) to following Dispatcher Subsystem. It includes: 1) menu or form interface to issue requests to the DSS; 2) customized visual (or graphical) forms to collect, revise customer/stock/company data. Additionally, the messages (of solution) emitted by the IDSS may consist of graphical presentation of some decisions; textual responses; displays of results via customized visual forms. For example, a classification with parameters (like certainty, weight, coefficients, etc) is depicted with the help of a 3-D graph or chart.

(2) **Investment Dispatcher:** Once receiving particular user requests, *investment dispatcher* will process them immediately. It may contain application server and gateways of database, model base and knowledge base. Note that this processing very well requires the *Investment Dispatcher* to draw on *Investment Provider* contents. The processing also changes the knowledge held in the *Investment Provider*. In either event, the *Investment Dispatcher* will issue a response to the user. Basically, the *Investment Dispatcher* has following abilities:

- **Knowledge acquisition:** for example, an *Investment Dispatcher* acquires knowledge about what a user wants the IDSS to do (i.e. request) or what is happening in the surrounding world (i.e. fundamental information, technical factors, risk information and market reaction, expert's opinion, etc). The *Investment Dispatcher* may draw on *Investment Provider* contents when using its acquisition ability. For example, its interpretation of a request (i.e., classification of positive, neutral or negative ranges of BHP at current time) may be based on previously acquired linguistic knowledge (i.e., *Turning Point or Ranges Models*). On the other hand, Knowledge acquisition ability can cause a change in the *Investment Provider*. For example, new knowledge obtained (i.e., company info, stock info, market info, etc) may be added in, edit or replace existing ones.
- **Knowledge selection/derivation:** *The Investment Dispatcher* is able to selectively recall or derives knowledge in following *Investment Provider* that forms a solution. For example, to classify BHP at current time, the *Investment dispatcher* may recall or derive knowledge from Experts' domain knowledge (i.e., experience of trading BHP stock), Fundamental knowledge (i.e., BHP company info), technical knowledge (i.e., Stock price and volume history), market reactions (i.e.,

contrarians' entry, market constituent) and so on. Additionally, it gives the ability to create, combine, revise, and delete models, knowledge and datasets in an *Investment Provider*. For example, some false or invalid market info (i.e., proven rumors) is identified useless and could be deleted from *Investment Provider*.

(3) **Investment Provider:** In response to *investment dispatcher's* command, the *Investment Provider* provides requested knowledge from model base, knowledge base, data base and expert domains (with fuzzy logic), which together will form a solution. It includes following components:

- **Model Bases:** It contains models and building blocks used to develop applications to run IDSS system. It may contain standard models (e.g. financial models like *DCF* or Risk models like *VaR*) and customized models (i.e. *Turning Point and Ranges Models*). Such models identify relationships among some collection of variables and numbers (i.e., constants). They may relate variables and constants within a series of numeric expressions involved in equations or inequalities, within a graph of nodes and arcs (e.g., a decision tree used in Figure 5 and 6), or via an arrangement of numbers in a data set. In effect, these are alternative ways to characterize a problem which a dispatcher uses to manufacture a solution. For example, using a *VaR* model to assess risks; using *ROI*, *IRR*, *NPV*, or a *Discounted Cash-flow* model to assess company value. All these models can be organized into what is called a model case. Note that problems that can be analyzed by the same dispatcher form a model class.
- **Knowledge Bases:** Knowledge may include descriptive, procedural and reasoning knowledge. For example, *Investment Provider* may contain knowledge about how long to retain historical data about a particular stock, or which inactive stock or invalid information should be discarded from the knowledge base. Additionally, descriptive knowledge will reside in the form of a particular company's data. A quantity of important knowledge comes from six modules discussed above. It may include (fuzzy) "IF-THEN" rules, for instance:
Rule 1:
IF: current P/E / history average of P/E < 0.8
 AND
 Current P/E / industry average < 0.7 AND
 Market value / company book value < 0.8 AND
 Market value / net asset value < 0.7 AND
 Company asset (land, plant, goodwill, etc) was undervalued by 30% AND
 Dividend payment is consistent in last 5 year
 AND
 Dividend payment is increasing by 10% each year AND

Cash flow is positive and healthy in last 3 years and future 1 year

THEN: Fundamental side of the stock is "GOOD" CF=80 (CF: Certainty Factor)

- **Data Bases:** Data could flow from several sources, i.e. Australian Centre for Advanced Computing & Communications (ac3) and other sources (i.e., announcements from ASX) and the databases formed contain variety of data forms and types. Their gateway provides necessary functions of data access, extraction, capture, integration and management. Particularly, it may be employed to accurately track and selectively recall descriptive knowledge in a relational database about a particular stock/company/market situation that satisfies a particular request or need.
- **Expert Domain Knowledge with Fuzzy Logic:** *Investment Provider* is also composed of expert (or other intelligent) domain knowledge and it provides the necessary execution and integration of the expert subsystems. It realizes that domain knowledge (or Investment sense) of experts (such as brokers, fund managers or large investors) could play an important role in IDSS execution process and make a real difference to sound investment decisions. For instance, the expert subsystem draws on descriptive and reasoning knowledge to infer advice in response to a user's request for a recommendation. The advice is comparable to that of a human expert. The expert knowledge acquisition process involves obtaining, structuring, and formalizing the knowledge of experts. Additionally, experts' domain knowledge and users' profiles could interactively affect IDSS (see Figure 1). For example, based on a broker's knowledge and experience (or investment sense), he could observe a unique *Contrarians' Entry Signal* and assess its importance. Then he can interact with IDSS system to manipulate above-discussed six modules to create a better result. On the other hand, results of individual modules could strengthen experts' knowledge base and then improve their interaction. Additionally, results of individual modules can become the basis for experts to decide which module to use or how to integrate them using particular optimization methods.

Note that often a human expert (i.e. a broker) is confronted with the task of making some sense out of a fuzzy situation, since one or more of the factors involved in the reasoning process is unclear or out of focus, in the sense that it has simultaneously several plausible appearances. For instance, an investment expert who reasons about the macroeconomic outlook (or company/stock outlook) might simultaneously

think that the outlook is good (60% certain) or fair (40% certain), each with some degree of certainty. The outlooks might be fuzzy. Therefore, fuzzy logic technology is useful in such fuzzy situation. Fuzzy variable and its own certainty factors become an important part of expert knowledge base.

IDSS may involve cooperation of a few of sub-expert systems which distribute in different sites (i.e. brokerage firms), thus it is necessary to organize them to integrate their intelligence and to achieve a better solution. For instance, stock broker A at *Sydney* can provide particular company info (i.e. BHP), stock broker B at *Melbourne* can provide a macroeconomic knowledge (i.e. energy), stock broker C at *Brisbane* can expertise in BHP stock traders' behaviors, and stock broker D at *Caines* can observe contrarian's activities. In order for an investor E at *Perth* to make a sound investment decision for BHP stock, it is necessary to integrate intelligence at five different sites. Therefore, I propose an agent-based distributed expert system which supports the sharing, managing, searching and presentation of distributed experts' knowledge. Note that such agent-based distributed expert systems have four kinds of cooperation which are *horizontal* cooperation, *tree* cooperation, *recursive* cooperation, and *hybrid* cooperation [18]. An intelligent agent is an object with its own knowledge and information base. Each intelligent agent acts in parallel with other intelligent agents and cooperates with a selected set of other agents to achieve a common set of goals. Intelligent agents must be able to recognize events, determine the meaning of those events and then take actions on behalf of a user. Figure 7 depicts sample *horizontal* agent-based distributed expert

systems, which may integrates usage of Matchmaker and Broker structure in different situations. For instance, if expert systems at Sydney, Melbourne and Perth involve a large volume of communication and communication speed is slow (type1), it is better to use Matchmaker structure, as once match was made, expert systems at Perth can communicate with ones at Sydney and Melbourne directly thus no need of matchmaker's involvement again. On the contrary, in case of expert systems at Brisbane, Caines and Perth, communication is quite fast and volume is small (Type 2), Broker structure is more suitable, as involvement of broker will not slow communication but will increase accuracy. Figure 8 depicts two types of distribution. In type 1, agent 2 of ES Melbourne & agent 3 of ES Sydney advertise their intelligence x (BHP company info) and y (energy sector's macroeconomic info) to Matchmaker (process 1a & 1b). Agent 1 of ES Perth needs intelligence of both x and y to make an investment decision, thus it asks matchmaker and the matchmaker answers with agent 2 and agent 3 (process 2 & 3). Once agent 1 of Perth receives the answer, it contacts agent 2 of ES Sydney and agent 3 of ES Melbourne directly and get corresponding intelligence x and y (process 4 & 5). In type 2, agent 1 of ES Perth can not contact agent 4 of Brisbane ES and agent 5 of Caines directly, on the contrary, it need to send request of p (BHP traders' behavior) and q (contrarian's observation) first to broker (process 6), then broker will get answers from agent 4 and 5 (process 7), then broker sends answers to agent 1 (process 8). Note in large distributed expert systems, both Matchmaker and Broker may be adopted, according to distribution difference and characteristics of expert systems.

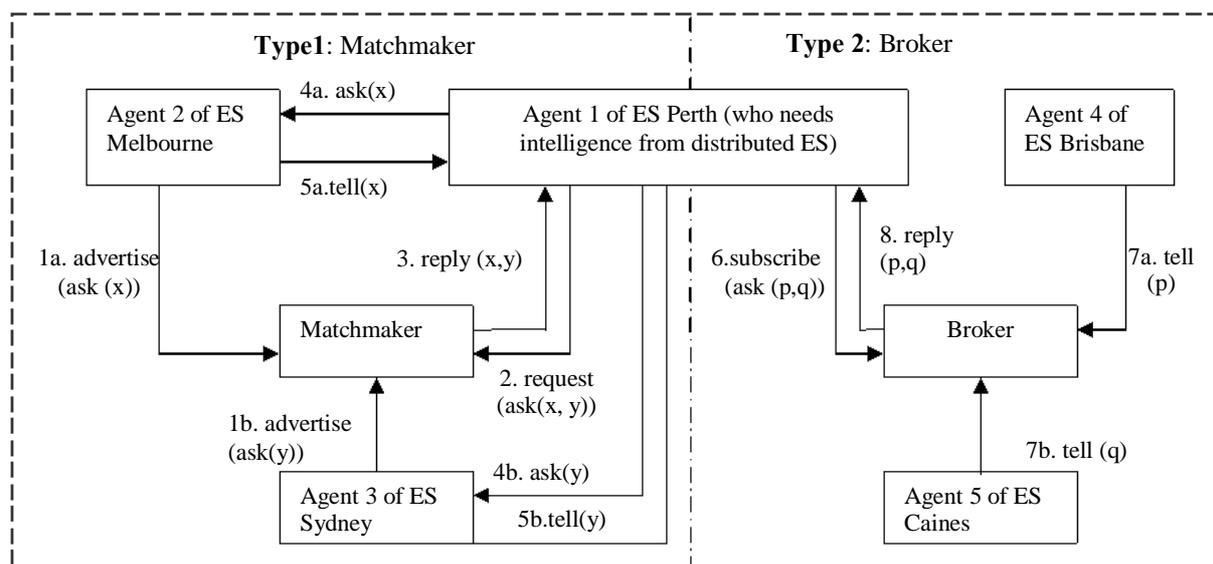


Figure 7: sample agent-based distributed expert systems

3. IDSS SYSTEM IMPLEMENTATION AND EVALUATION

Proposed IDSS is a KB-DSS in essence, thus building it involves the capabilities, functionality and structures of both DSS and ES with the emphasis on the support of DSS. I adopted a KB-DSS methodology proposed by Klein and Methlie [19]. Additionally, Evolutionary Prototyping proposed by Turban and Aronson [20] was used as an aid. The initial prototype was implemented by mainly using Jess and Java. An industry partner, Tricom (www.tricom.com.au), contributed expert domain knowledge and involved in IDSS implementation.

Based on initial prototype, I got following initial evaluation results:

- 1) Real transaction results. For instance, IDSS system picks up 10 stocks at an experiment and return is over 20%, which is higher than benchmark (i.e. ASX index) and most fund performance.
- 2) Experts' feedback. Experts (such as brokers, fund managers or large investors) are users and interact with the system. An industry partner TRICOM (broker) gave positive feedbacks (like ease of use, usefulness or other features) in both experiment and real trading practice.
- 3) Compared to other stock trading support systems, IDSS shows better performance in experiments, including effectiveness, ease of use, powerfulness of system functions, efficiency, etc.

While further implementation is in progress and more evaluation will be done, I expect a more efficient IDSS to be finalized and more solid experimental results to be obtained by December 2004.

4. CONCLUSIONS

I propose an Investment Decision Support System (IDSS) in order to get better mid-term investment results than conventional methods, through identification and classification of Positive, Neutral and Negative investment opportunity ranges and related risks. This IDSS system takes advantage of and optimizes existing methods and newly proposed methods - *Turning Point Model* (in finance) and *Optimized AutoSplit method* (in IT), which can help better identify investment opportunities and risks features. Initial system evaluation shows that the IDSS system (with proposed methods) can successfully identify and classify investment opportunity ranges and other hidden variables, outperforming the benchmark index and other systems.

REFERENCE

[1] O'Hara, M., *Market Microstructure Theory*. Cambridge: Basil Blackwell, 1995.

- [2] Fama, E., French k., "Size and book-to-market factors in earnings and returns," *Journal of Finance*, 50: 131-155, 1995
- [3] La Porta R., Lakonishok, J., Shliefer, A., R. Vishny., "Good news for value stocks: Further evidence on market efficiency", *Journal of Finance*, 52:859-874, 1997.
- [4] Joffe B., "Fundamental Analysis", <http://www.sharenet.co.za/free/library/fundamen.htm>, accessed on 11/08/2004
- [5] Pike, T., "Technical Analysis", <http://www.sharenet.co.za/free/library/technica.htm>, accessed on 11/08/2004.
- [6] Mendelsohn L.B., "Artificial Intelligence in Capital Markets, Chapter 5: Global Trading Utilizing Neural Networks: A Synergistic Approach", *Virtual Trading*, Probus Publishing Company, Chicago, Illinois, 1995
- [7] Allen H, Taylor P, "Chart analysis and the foreign exchange market". *Bank of England Bulletin*, November: 547-51, 1989.
- [8] Carter R.B., Van Auken H.E. "Security analysis and portfolio management: a survey and analysis". *Journal of Portfolio Management*, 16: 81-85, 1990.
- [9] Gencay, R., "The predictability of security returns with simple technical trading rules". *Journal of Empirical Finance*, 5: 347-359, 1998
- [10] Kohara, K., Ishikawa, T., Fukuhara, Y., Nakamura, Y., "Stock price prediction using prior knowledge and neural networks". *Intelligent Systems in Accounting, Finance and management*, 6:11-22, 1997.
- [11] Lee, K.M., Kim, W. X., "Integration of human knowledge and machine knowledge by using fuzzy post adjustment: its performance in stock market timing prediction". *Expert Systems*, 12 (4), 331-338, 1995
- [12] Kim, k., Han, I., "The extraction of trading rules form stock market data using rough sets". *Expert Systems*, 18 (4), 194-202
- [13] Last, M., Klein, Y., Kandel, A., *Knowledge discovery in time series databases*. IEEE Transactions on Systems, Man, and Cybernetics-PartB: Cybernetics, 31 (1), 160-169, 2001
- [14] Wang, Y., "Predicting stock price using fuzzy grey prediction system". *Expert Systems with Applications*, 22(1), 33-38, 2001
- [15] Armstrong, J., Collopy, F.k., "Causal forces: structuring knowledge for time series extrapolation". *Journal of Forecasting*, 12: 103-115, 1993.
- [16] Skouras, S., "Financial returns and efficiency as seen by an artificial technical analyst". *Journal of Economic Dynamics and Control*, 25, 213-244, 2001.
- [17] Pan J., Kitagawa H., Faloutsos C., Hamamoto M., "AutoSplit: Fast and Scalable Discovery of Hidden Variables in Stream and Multimedia Databases". *Proceedings of the The Eighth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2004)*, Australia, May 26~28, 2004
- [18] Zhang, C., "Cooperation under uncertainty in distributed expert systems", *Artificial Intelligence*, 56: 21-69, 1992
- [19] Klein, M.R., Methlie, L.B., *Knowledge-based decision support systems with Applications in Business*, 2nd Edition. England: John Wiley & Sons, 1995.
- [20] Turban E., Aronson J.E., *Decision Support Systems and Intelligent Systems*, Prentice Hall, Upper Saddle River, NJ, 1998

Bayesian Network Induction with Incomplete Private Data

Justin Zhan¹, LiWu Chang², Stan Matwin³

^{1,3} School of Information Technology & Engineering, University of Ottawa, Canada

² Center For High Assurance Computer Systems, Naval Research Laboratory, USA
¹³{zhizhan, stan}@site.uottawa.ca, ²lchang@itd.nrl.navy.mil

ABSTRACT

A Bayesian network is a graphical model for representing probabilistic relationships among a set of variables. It is an important model for business analysis. Bayesian network learning methods have been applied to business analysis where data privacy is not considered. However, how to learn a Bayesian network over private data presents a much greater challenge. In this paper, we develop an approach to tackle the problem of Bayesian network induction on private data which may contain missing values. The basic idea of our proposed approach is that we combine randomization technique with Expectation Maximization (EM) algorithm. The purpose of using randomization is to disguise the raw data. EM algorithm is applied for missing values in the private data set. We also present a method to conduct Bayesian network construction, which is one of data mining computations, from the disguised data.

Keywords: Bayesian network, randomization, EM algorithm, privacy.

1. INTRODUCTION

A Bayesian network is a high-level representation of a probability distribution over a set of variables that are used for constructing a model of the problem domain. It has been widely used in sales decision making, marketing systems, risk analysis, cost benefit factor inference in E-services, and other business applications. As Bill Gates mentioned, in Los Angeles Times on October 28, 1996, that Microsoft's competitive advantage was its expertise in Bayesian network. The Bayesian framework offers many advantages over alternative modeling approaches. A Bayesian network can be used to compute the predictive distribution on effects of possible actions since it is a model of the problem domain probability distribution. For instance, it can be used for the optimal procedure of making decision in risk analysis. Bayesian network model has been also found to be very robust to tolerate small alterations. Thus, it can be used for sales marketing systems since these systems need to be able to follow market changes rapidly without making much modification on the model. In addition, expert domain knowledge can be coded as prior distribution in Bayesian modeling.

Over the last decade, the Bayesian network has been widely utilized and various techniques have been developed to learn Bayesian networks from data. Although the techniques that have been developed are effective, new techniques dealing with Bayesian network induction over private data are required. In other words, we need methods to learn a Bayesian network with data privacy preserved. In particular, we try to solve the following type of problem: To improve their services, business agents want to obtain certain information about their customers. For instance, the information of interest can be the probability that a customer buys butter given that she bought bread. In

order to collect the information from its customers, a business agent sends out a survey containing a set of questions. Customers are expected to answer those questions and send their answers back. However, because the survey contains questions regarding private information, not every user feels comfortable to disclose her answers to those questions. To protect customers' privacy, they need somehow mask their answers before sending them back to the agent. We observe that randomization can be used to hide customers' answers. This is because on one hand the raw data values of the original customer's information can be disguised via randomization, on the other hand, their probability distribution can be approximately estimated from the disguised data by agents for business analysis. In practice, even though randomization protection techniques are utilized, some customers still do not provide their data. Therefore, the collected data usually contain missing values. The challenge is how to conduct data mining from the disguised data with missing values?

To address this challenge, we propose to combine the proposed *Multi-Group Randomization* technique with the Expectation Maximization (EM) algorithm [2]. The basic idea of multi-group randomization technique is to partition attributes of a data set into several groups and randomize data in each group independently so that business agents can't tell with probabilities better than a pre-defended threshold whether the data from a customer contain truthful information or false information. Although information from each individual customer is scrambled, if the number of customers is significantly large, the aggregate information of these customers can be estimated with decent accuracy. Such property is useful for building a Bayesian network, since it is based on aggregate values of a data set, rather than individual data items. The EM algorithm is a general iterative algorithm for parameter estimation by

maximum likelihood when some of the random variables involved are not completely observed. It formalizes an intuitive idea for obtaining parameter estimates when some of the data are missing. The term EM was introduced in [2] where proof of general results about the behavior of the algorithm was given as well as a large number of applications was introduced.

The rest of the paper is organized as the follows: We describe the background of Bayesian network in Section 2. In Section 3, we briefly describe how multi-group randomized response technique works. Then in section 4, we describe the EM algorithm. We give an algorithm to build a Bayesian network on disguised data with missing values in Section 5. In Section 6, we discuss related work. We give our conclusion in Section 7.

2. BACKGROUND OF BAYESIAN NETWORKS

A Bayesian network [10] is a representation of probabilistic conditional independence. The network model is a directed acyclic graph (DAC) in which a node represents a random variable and a directed link or an arrow denotes the conditional probability of a child node given its parent node with the arrow pointing from the parent to the child. Suppose a network BN_1 have n nodes. Associated with each node (e.g., X_i , $i \in [1, \dots, n]$) is a conditional probability table (CPT) that shows the probabilistic dependency relationships between X_i and its parent nodes, $Parent(X_i)$, i.e., $P(X_i = x_i | Parent(X_i) = pa_i, D, BN_1)$, where D is the dataset, BN_1 is the network model, and x_i and pa_i are values of X_i and $Parent(X_i)$, respectively. An entry of the CPT is a network parameter.

A Bayesian network may be inductively learnt from a data set [6]. As discussed in [6], inductive Bayesian learning involves the selection of network models and the estimation of parameters (e.g., the CPTs). Given a complete data set D , learning of the best Bayesian network model is carried out by computing the posterior probability $P(BN_1 | D)$ for different BN_1 , and selecting the one with the largest value. From the Bayes rule, $P(BN_1 | D)$ is determined by the likelihood $P(D | BN_1)$, the prior $P(BN_1)$, and the data probability D :

$$P(BN_1 | D) = \frac{P(D | BN_1) \cdot P(BN_1)}{P(D)}$$

Given the prior $P(BN_1)$, the model selection is determined by $P(D | BN_1)$. Assume the best model is denoted as BN . The network parameters are estimated with the data of D and the prior probability distribution of parameters. In this paper, we assume the Bayesian

network model is known, and will mainly consider the induction of parameters or the conditional probability tables. Suppose the uniform prior is used. (See [6] for different prior probabilities.) The CPT entry associated with the values of $X_i = x_i$ and $Parent(X_i) = pa_i$ is as follows:

$$P(X_i = x_i | Parent(X_i) = pa_i, D, BN_1) = \frac{\#(X_i = x_i, Parent(X_i) = pa_i) + 1}{\#(Parent(X_i) = pa_i) + |X_i|} \quad \text{EQ.(1)}$$

for a complete data set D (i.e., D without missing value), where $\#(\cdot)$ stands for the total number of the records that satisfies conditions in (\cdot) , and $|X_i|$ denotes the number of values that X_i has. Divide both the numerator and the denominator of EQ.(1) by the total number of data records ($|D|$) to obtain

$$\frac{\#(X_i = x_i, Parent(X_i) = pa_i) / |D| + 1 / |D|}{\#(Parent(X_i) = pa_i) / |D| + |X_i| / |D|} = \frac{P(X_i = x_i, Parent(X_i) = pa_i) + 1 / |D|}{P(Parent(X_i) = pa_i) + |X_i| / |D|}$$

In this paper, we will use $P(X_i = x_i)$ to denote the probability measure obtained from the frequency count $\#(X_i = x_i) / |D|$.

3. MULTI-GROUP RANDOMIZED RESPONSE TECHNIQUES

Randomized Response techniques were firstly introduced in [11] as a technique to solve the following survey problem: to estimate the percentage of people in a population that has attribute A , queries are sent to a group of people. Since the attribute A is related to some confidential aspects of human life, respondents may decide not to reply at all or to reply with incorrect answers. To enhance the level of cooperation, instead of asking each respondent whether she has attribute A , the interviewer asks each respondent two related questions, the answers to which are opposite to each other [11]. For example, the questions could be like the following. If the statement is correct, the respondent answers "yes"; otherwise, she answers "no".

1. I have the sensitive attribute A .
2. I do not have the sensitive attribute A .

Respondents use a randomizing device to decide which question to answer, without letting the interviewer know which question is answered. The randomizing device is designed in such a way that the probability of choosing the first question is q , and the probability of choosing the second question is $1 - q$. Although the interviewer learns the responses (e.g., "yes" or "no"), he does not know which question was answered by the respondents. Thus the respondents' privacy is preserved. Since the interviewer's interest is to get the answer to the first

question, and the answer to the second question is exactly the opposite to the answer for the first one, if the respondent chooses to answer the first question, we say that she is telling the truth; if the respondent chooses to answer the second question, we say that she is telling a lie. To estimate the percentage of people who has the attribute A, we have

$$P^*(A = yes) = P(A = yes) \cdot q + P(A = no) \cdot (1 - q)$$

$$P^*(A = no) = P(A = no) \cdot q + P(A = yes) \cdot (1 - q)$$

where $P^*(A = yes)$ is the proportion of the "yes" responses obtained from the survey data. $P^*(A = no)$ is the proportion of the "no" responses obtained from the survey data. $P(A = yes)$ is the estimated proportion of the "yes" responses to the sensitive questions. $P(A = no)$ is the estimated proportion of the "no" responses to the sensitive questions. Getting $P(A = yes)$ and $P(A = no)$ is the goal of the survey. By solving the above equations, we can get $P(A = yes)$ and $P(A = no)$ if $q \neq 0.5$.

In this section, we develop a multi-group randomized response technique to deal with the problem of Bayesian network construction where multiple attributes are normally involved.

3.1 Notations

In this paper, we assume data are binary. Suppose there are N attributes (A_1, \dots, A_N) in a data set. Let L represent any logical expression based on those attributes (e.g., $L = (A_1 = 1) \wedge (A_2 = 0)$); let \bar{L} denote the logical expression that reverses the 1's in L to 0's and 0's to 1's; we call \bar{L} the opposite of L in value assignments. For example, for the L in the previous example, $\bar{L} = (A_1 = 0 \wedge A_2 = 1)$. Let $P^*(L)$ be the proportion of the records in the whole *disguised* data set that satisfy $L = true$. Let $P(L)$ be the proportion of the records in the whole *undisguised* data set that satisfy $L = true$ (the undisguised data set contains the true data, but it does not exist). $P^*(L)$ can be observed directly from the disguised data, but $P(L)$, the actual proportion that we are interested in, cannot be observed from the disguised data because the undisguised data set is not available to anybody; we have to estimate $P(L)$. Our goal is to find a way to estimate $P(L)$ from $P^*(L)$. In our multi-group scheme, we also divide each expression L into multiple sub-expressions. For example, in a m-group scheme, we write $L = L_1 L_2 L_3 \dots L_m$, where L_i contains only the attributes in the group i.

3.2 One-Group Scheme

In the one-group scheme, all the attributes are put in the same group, and all the attributes are either reversed

together or keeping the same values. In other words, when sending the private data to the central database, users either tell the truth about all their answers to the sensitive questions or tell the lie about all their answers.

The probability for the first event is q , and the probability for the second event is $1 - q$. The general model for the one-group scheme is described in the following:

$$P^*(L) = P(L) \cdot q + P(\bar{L}) \cdot (1 - q)$$

$$P^*(\bar{L}) = P(\bar{L}) \cdot q + P(L) \cdot (1 - q)$$

L denotes any logic expression among the attributes, for instance, L could be $(A_1 = 0, A_2 = 1, A_3 = 0, A_4 = 0)$ and represents the case when attribute 1 is 0, attribute 2 is 1, attribute 3 is 0, and attribute 4 is 0. The above model can be simplified. Let M_1 denote the coefficients matrix of the above equations, and let $p = q$ and $q = (1 - q)$, then

$$\begin{Bmatrix} P^*(0Bar) \\ P^*(1Bar) \end{Bmatrix} = M_1 \begin{Bmatrix} P(0Bar) \\ P(1Bar) \end{Bmatrix} \quad \text{Eq. (3)}$$

Where $0Bar = L$, $1Bar = \bar{L}$, and

$$M_1 = \begin{bmatrix} p & q \\ q & p \end{bmatrix}$$

3.3 Two-Group Scheme

In the one-group scheme, if the interviewer somehow knows whether the respondents tell a truth or a lie for one attribute, she can immediately obtain all the true values for all other attributes of a respondent's response. To improve the privacy level, data providers divide all the attributes into two groups (All the data providers should group the attributes in the same way, e.g., if one user lets attribute A_1 and A_2 be in group 1, then other users also let attribute A_1 and A_2 be in group 1). They then apply the randomized response techniques for each group *independently*. For example, the users can tell the truth for one group while telling the lie for the other group. With this scheme, even if the interviewers know information about one group, they will not be able to derive the information for the other group because they are disguised independently.

To show how to estimate $P(L_1 L_2)$, we look at all the contributions to $P^*(L_1 L_2)$. There are four parts that contribute to $P^*(L_1 L_2)$:

1. $P(L_1 L_2)$: users tell the truth about all the answers for both groups; the probability for this event is q^2 .

2. $P(L_1 \bar{L}_2)$: users tell the truth about all the answers for group 1 and tell the lie about all the answers for group 2; the probability for this event is $q \cdot (1-q)$.

3. $P(\bar{L}_1 L_2)$: users tell the lie about all the answers for group 1 and tell the truth about all the answers for group 2; the probability for this event is $q \cdot (1-q)$.

4. $P(\bar{L}_1 \bar{L}_2)$: users tell the lie about all the answers for both groups; the probability of this event is $(1-q)^2$.

We then have the following equation:

$$P^*(L_1 L_2) = P(L_1 L_2) \cdot q^2 + P(L_1 \bar{L}_2) \cdot q(1-q) + P(\bar{L}_1 L_2) \cdot q(1-q) + P(\bar{L}_1 \bar{L}_2) \cdot (1-q)^2$$

There are four unknown variables in the above equation ($P(L_1 L_2), P(L_1 \bar{L}_2), P(\bar{L}_1 L_2), P(\bar{L}_1 \bar{L}_2)$). To solve the above equation, we need three more equations. We can derive them using the similar method. The final equations are described in the following:

$$\begin{Bmatrix} P^*(0Bar) \\ P^*(1Bar) \\ P^*(2Bar) \end{Bmatrix} = M_2 \cdot \begin{Bmatrix} P(0Bar) \\ P(1Bar) \\ P(2Bar) \end{Bmatrix}$$

where $0Bar = L_1 L_2, 1Bar = L_1 \bar{L}_2 + \bar{L}_1 L_2, 2Bar = \bar{L}_1 \bar{L}_2, M_2$ is the coefficients matrix, and let $p = q$ and $q = 1-q$, then,

$$M_2 = \begin{bmatrix} p^2 & pq & q^2 \\ 2pq & p^2 + q^2 & 2pq \\ q^2 & pq & p^2 \end{bmatrix}$$

By solving the above equations, we can get $P(L_1 L_2) = P(0Bar)$

3.5 Multi-Group Scheme

Similar techniques can be employed to extend the above schemes to three-group scheme, four-group scheme, and so on. In the following, we will give a general formula for the m-group scheme, where we apply randomized response techniques for each group *independently*.

$$\begin{Bmatrix} P^*(0Bar) \\ P^*(1Bar) \\ \vdots \\ P^*(mBar) \end{Bmatrix} = M_m \cdot \begin{Bmatrix} P(0Bar) \\ P(1Bar) \\ \vdots \\ P(mBar) \end{Bmatrix}$$

Where

$0Bar = L_1 L_2 L_3 \dots L_m, 1Bar = \bar{L}_1 L_2 L_3 \dots L_m, \dots, mBar = \bar{L}_1 \bar{L}_2 \bar{L}_3 \dots \bar{L}_m$ and M_m is the coefficients matrix. Let $p = q$ and $q = 1 - q$, then

$$M_m = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1(n+1)} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2(n+1)} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3(n+1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{(n+1)1} & a_{(n+1)2} & a_{(n+1)3} & \dots & a_{(n+1)(n+1)} \end{bmatrix}$$

The coefficient matrix is different for different group scheme. The values of a_{ij} can be derived as we did for two-group scheme. After we derive the coefficient matrix, we can solve the above equation and obtain $P(E_1 E_2 \dots E_n) = P(0Bar)$ since $P^*(0Bar), P^*(1Bar), \dots, P^*(mBar)$, can be obtained from the randomized data.

4. EM ALGORITHM

The Expectation Maximization (EM) algorithm [2, 8] is a general algorithm for maximum likelihood estimation where the data are incomplete or the likelihood function involves hidden variables. The EM algorithm starts with randomly assigned values to all the parameters to be estimated. It then iteratively alternates between two steps, called the expectation step (i.e., the E-step) and the maximization step (i.e., the M-step), respectively. In the E-step, it computes the expected likelihood for the complete data (Q-function) where the expectation is taken with regards to the computed conditional distribution of the hidden variables given the current settings of parameters and our observed data. In the M-step, it re-estimates all the parameters by maximizing the Q-function. Once we have a new generation of parameter values, we can repeat the E-step and M-step. The above process continues until the likelihood converges.

Assume that the joint probability for hidden data Y and observed data Z is parameterized using m , as $P(y, z | m)$. The marginal probability for Z is then $P(z | m) = \sum_y P(y, z | m)$. Given observed data z, we

want to find the value of m that maximizes the log likelihood, $L(m) = \log P(z | m)$. The procedures of the EM algorithm are as follows:

1. Initialize $m^{(0)}$ randomly or heuristically according to prior knowledge about what the optimal parameter value might be.
2. Iteratively improve the estimate of m by alternating between the following two steps:
 - (a) E-step: Compute a distribution $\hat{P}^{(i)}$ over the range Y

such that $\hat{P}^{(t)} = P(y | z, m^{t-1})$.

(b) M-step: Set $m^{(t)}$ to the m that maximizes $E_{\tilde{p}}(\log P(y, z | m))$.

3. Stop when the algorithm converges to a local maximum.

5. CONSTRUCTION OF A BAYESIAN NETWORK

The problem of learning Bayesian networks comes in several varieties. The structure of the network can be either *known or unknown*, and the variables in the network can be either *fully observable or with missing values*. In this paper, we consider the case of *known structure and with missing values*. Since the structure is known, construction requires only the computation of the CPT for each node. We will describe how to compute the CPTs when data set is complete, and then present an algorithm to deal with Bayesian network construction when the data set D is disguised and incomplete. Let parent nodes of the i th node be $Y_{1,L}, Y_p$. To obtain CPT entries of node i , we need to compute the two terms:

$$P(X_i = x_i, Y_{1,L} = y_{1,L}, Y_p = y_p) \text{ and } P(Y_{1,L} = y_{1,L}, Y_p = y_p)$$

Since

$$P(X_i = x_i | Parent(X_i) = pa_i, D, BN) = \frac{P(X_i = x_i, Y_{1,L} = y_{1,L}, Y_p = y_p) + 1/|D|}{P(Y_{1,L} = y_{1,L}, Y_p = y_p) + |X_i|/|D|}$$

When data set D is undisguised and complete, these two terms can be easily computed. But, it is not the case if D is disguised and incomplete. We provide the following algorithm to deal with Bayesian network construction when the data set D is disguised and incomplete.

Step I: Estimate missing values

Use EM algorithm to estimate the missing values in the disguised data set and obtain a complete data set CD. Please note that the distribution of randomized data very likely differs from the data distribution before randomization. Therefore, we need test the data distribution after randomization and apply proper EM algorithm.

Step II: Compute CPTs

Apply the estimation model of the randomized technique described in Section 3 to estimate those terms needed for the CPT entries. Consider the Wet-grass example, where wet-grass (W) can be caused by either rain (R) or sprinkling (S). The particular CTP of grass being wet, given rain and sprinkling is as follows.

$$P(W = t | S = t, R = t, CD, BN) = \frac{P(W = t, S = t, R = t) + 1/|CD|}{P(S = t, R = t) + |W|/|CD|}$$

(|CD| denotes the total number of records in the complete data set.) Assume the data set is disguised using one-group scheme. We compute $P^*(W = t, S = t, R = t)$ and $P^*(W = f, S = f, R = f)$ on the data set CD. We then apply one-group estimation model to compute $P(W = t, S = t, R = t)$.

$$\begin{Bmatrix} P^*(W = t, S = t, R = t) \\ P^*(W = f, S = f, R = f) \end{Bmatrix} = M_1 \cdot \begin{Bmatrix} P(W = t, S = t, R = t) \\ P(W = f, S = f, R = f) \end{Bmatrix}$$

where $M_1 = \begin{bmatrix} q & 1-q \\ 1-q & q \end{bmatrix}$. Similarly, $P(S = t, R = t)$ can be

computed. Hence, we can calculate $P(W = t | S = t, R = t)$. With a similar calculation, other CPTs can be obtained.

6. PREVIOUS WORK

There are two existing approaches to solve privacy preserving data mining problems. One is the perturbation approach, the other is the secure multi-party computation approach. Agrawal and Srikant proposed a scheme for privacy-preserving data mining using random perturbation [3]. In their scheme, a random number is added to the value of a sensitive attribute. For example, if x_i is the value of a sensitive attribute, $x_i + r$, rather than x_i , will appear in the database, where r is a random value drawn from some distribution. The paper shows that if the random number is generated with some known distribution (e.g., uniform or Gaussian distribution), it is possible to recover the distribution of the values of that sensitive attribute. With the independence assumption of the attributes, the paper then shows that a decision tree classifier can be built with the knowledge of distribution of each attribute. Du and Zhan [4] proposed to use randomized response techniques [11] to tackle privacy-preserving data mining problem. In their scheme, they do not assume the independence of the attributes. Rizvi and Haritsa presented a scheme called MASK to mine associations with secrecy constraints in [9], and Evfimievski et al. proposed an approach to conduct privacy preserving association rule mining based on randomization techniques [1]. Kargupta et al. further analyze the effectiveness of randomization approach in [5]. Several SMC-based privacy preserving data mining schemes have been proposed [7, 7, 13]. These studies mainly focused on two-party distributed computing, and each party usually contributes a set of records. In our proposed research, we focus on centralized computing, and each participant only has certain number of records to contribute. All records are

combined together into a central database before the computation begins.

7. CONCLUSION AND FUTURE WORK

As mentioned, Bayesian network learning is proven to be valuable to the business world. In this paper, we have combined multi-group randomized response techniques with EM algorithm for constructing Bayesian networks over disguised data set where some data values are missing. In our future work, we will continue refining our technique to handle various risk-analyses in business. We will also extend our solution to data sets consisting of non-binary data types.

REFERENCES

- [1] J. Gehrke, A. Evfimievshi, and R. Srikant, "Limiting privacy breaches in privacy preserving data mining", In proceedings of the 22nd ACM SIGACT-SIGMOD -SIGART symposium on principles of database systems, San Diego, CA, June, 2003.
- [2] N. Laird, A. Dempster, D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm, the royal statistical society, Series B, Methodological, 39:1-38, 1977.
- [3] W. Barksdale, "New randomized response techniques for control of non-sampling errors in surveys", PhD thesis, University of North Carolina, Chapel Hill, 1971.
- [4] W. Du, Z. Zhan, "Using randomized response techniques for privacy-preserving data mining", In proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining, Washington, DC, USA, August 24-27, 2003.
- [5] Q. Wang, H. Kargupta, S. Datta, K. Sivakumar, "Random data perturbation techniques and privacy preserving data mining", In the IEEE international conference on data mining, Florida, USA, 2003.
- [6] D. Heckerman, "Bayesian networks for knowledge discovery", In Advances in knowledge discovery and data mining, AAAI Press/MIT Press, 1996.
- [7] Y. Lindell and B. Pinkas, "Privacy preserving data mining", Advances in cryptology-CRYPTO'00, 1880 of lecture notes in computer science, Spinger-Verlag: 36-54, 2000.
- [8] R. Neal, G. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants", Learning in graphical models, editor: M. Jordan, 1998.
- [9] S. Rizvi, J. Haritsa, "Maintaining data privacy in association rule mining", In proceedings of the 28th VLDB conference, Hong Kong, China, 2002.
- [10] S. Russell, P. Norvig, "Artificial Intelligence: A modern approach", Prentice Hall, Upper Saddle River, New Jersey 07458, 1995.
- [11] S. Warner, "Randomized response: A survey technique for eliminating evasive answer bias", Journal of American Statistical Association, 60(309):63-69, March, 1965.
- [12] A. Tamhane, "Randomized response techniques for multiple sensitive attributes", the American Statistical Association, 76(376):916-923, December, 1981.
- [13] J. Vaidya, C. Clifton, "Privacy preserving association rule mining in vertically partitioned data", In proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining, July 23-26, 2002.

Constructing Control Process for Wafer Defects Using Data Mining Technique

Leeing Tong, Hsingyin Lee, Chifeng Huang, Changke Lin, Chienhui Yang

Department of Industrial Engineering and Management,
National Chiao Tung University, Hsinchu, Taiwan, China

litong@cc.nctu.edu.tw, {chaba.iem89, frogman.iem89, changke.iem89, mimosa.iem90g}@nctu.edu.tw

ABSTRACT

The wafer defects influence the yield of a wafer. The integrated circuits (IC) manufacturers usually use a Poisson distribution based c-chart to monitor the lot-to-lot wafer defects. As the wafer size increases, defects on wafer tend to cluster. When the c-chart is used, the clustered defects frequently cause erroneous results. The main objective of this study is to develop a hierarchical adaptive control process to monitor the clustered defects effectively and detect the wafer-to-wafer variation and lot-to-lot variation simultaneously using data mining technique.

Keywords: wafer defects, c-chart, defect clustering, adaptive control chart, data mining.

1. INTRODUCTION

The IC fabrication has been the most popular industry in recent decades. The key competitive ability of IC products is the wafer yield. The wafer yield is defined by the percentage of the chip without defects on a wafer. To enhance the wafer yield, most IC manufacturers employ a Poisson distribution based defect control chart (c-chart) to monitor the wafer defects. The c-chart is proved to be effective when wafer size is small (e.g., four or six inches wafers). As the wafer size has been enlarged to twelve-inch, the manufacturing process becomes very complicated. Consequently, the clustering phenomenon of the defects becomes increasingly apparent. This phenomenon will affect the accuracy of the c-chart. When the conventional c-chart is used to monitor the large-size wafers, the clustered defects frequently cause erroneous results. That is, the c-chart may cause too many false alarms. This is because c-chart requires that the occurrence of a defect in any location is independent of the occurrence of defects in other locations.

The IC manufacturing process is composed of hundreds of steps. At each step, many possible factors may affect the quality of a wafer. There are three types of wafer variations: within-wafer variation, wafer-to-wafer variation and lot-to-lot variation. The c-chart can only detect the lot-to-lot variation under the condition that the defects are scattered randomly on a wafer and there is no wafer-to-wafer variation. Besides, the cost of sampling wafers from a lot becomes expensive for large-size wafers. An adaptive control chart may be utilized to reduce the sampling cost. Therefore, the main objective of this study is to develop a complete adaptive detecting procedure of wafer defects, which concerns both multiple variations and clustering phenomenon. Using the proposed procedure, the wafer defects can be controlled efficiently with low sampling cost.

2. DOCUMENTATION DISCUSSION

2.1 Clustering analysis [4]

Cluster analysis is a popular data mining technique used for combining observations into groups or clusters such that each cluster is homogeneous with respect to certain characteristics and observations of one cluster should be different from the observations of other clusters. In hierarchical cluster analysis, clusters are formed hierarchically such that the number of clusters at each step is $n-1$, $n-2$, $n-3$, and so on. A number of different algorithms for hierarchical clustering were developed. These algorithms differed mainly with respect to how distances between two clusters are computed. The Euclidean distance for p variables is a common formula used for computing the distance between two observations. The distance is utilized as a measure of similarity in cluster analysis.

2.2 Three-step hierarchical control process

In the IC manufacturing process, there are three types of wafer variations: within-wafer variation, wafer-to-wafer variation and lot-to-lot variation may cause the yield loss. Wells and Smith [5] claimed that the three variations must be under control to assure the quality of a wafer and they proposed a step-by-step control process as follows:

(1) Determine the uniformity of a wafer.

Calculate the non-uniformity (NU) of each wafer:

$$NU = 2V \text{ and } V = 100\hat{s} / \bar{X} \quad (1)$$

Where \hat{s} and \bar{X} represents the estimated standard deviation and mean.

(2) Draw $\bar{X}_{NU} - R_{NU}$ control chart using the values of NU to determine whether the non-uniformity of each wafer is under control.

- (3) Determine the lot-to-lot variation.
 - (i) Calculate the average run value of every lot.
 - (ii) Draw $\bar{X}-MR$ chart to determine whether the lot-to-lot variation is under control.

(4) Calculate the total variations of the process

$$\hat{s}_{TOTAL} = (\hat{s}_R^2 + \hat{s}_W^2 + \hat{s}_L^2)^{1/2} \quad (2)$$

The estimated within wafer variation can be calculated using the average non-uniformity of every lot. The formula is given as follows:

$$\hat{s}_L = (\overline{X_{NU}})(\overline{X})/200 \quad (3)$$

Moreover, the wafer-to-wafer variation can also be obtained using the average range of every lot. The formula is given as follows :

$$\hat{s}_W = \overline{R}/d_2 \quad (4)$$

Finally, the estimated lot-to-lot variation can be obtained using the average moving-range of every lot. The formula is given as follows :

$$\hat{s}_R = \overline{MR}/d_2 \quad (5)$$

The control process proposed by Wells and Smith has two drawbacks: it can only detect the large shift of the process, that is, it is not sensitive to the small shift; it can only be used in the quantitative quality characters.

2.3 Adaptive control chart

Adaptive control chart is a special kind of quality control tool in which the sample size and the sampling interval are changeable. It utilizes the previous sample data as a basis to fluctuate the sampling sizes and sampling intervals. When sample points are near out of control, a stricter sampling scheme (e.g., larger sample size and shorter sampling period) is adopted to detect the out-of control situation efficiently. When sample points are near the target value, a looser sampling scheme (e.g., smaller sample size and longer sampling period) is adopted to save the sampling cost.

In order to determine when to adopt the strict, normal or loose sampling scheme, the upper and lower warning limits must be constructed within the upper and lower control limits. If a sample point falls within the warning limits, a loose sampling scheme is utilized. If a sample point falls between the warning limits and the control limits, a strict sampling scheme should be utilized. The basic model of an adaptive control chart is depicted in Fig.1.

2.4 \bar{X} adaptive control chart [3]

The assumptions and parameters of designing a \bar{X} adaptive control chart can be described as follows:

- (1) Assumptions
 - (i) The observations follow a Normal

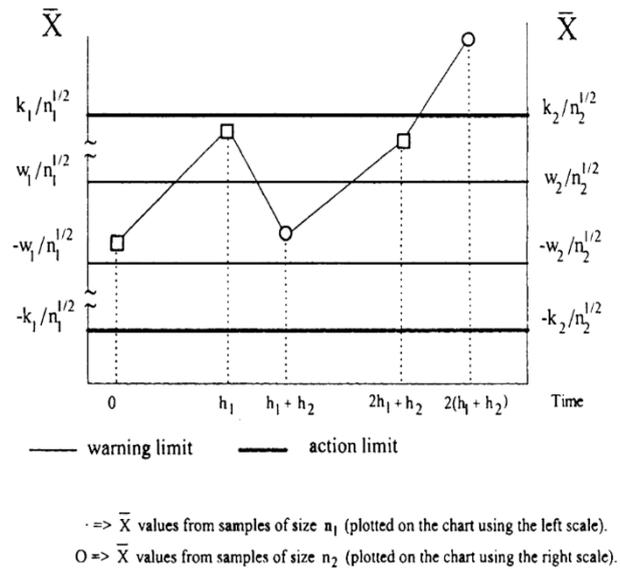


Figure.1 Adaptive control chart

distribution.

- (ii) Suppose the sample size and the sampling interval of the traditional control chart are n_0 and t_0 , respectively. A sampling scheme of n_1 (small sample size) and h_1 (long sample interval), or n_2 (large sample size) and h_2 (short sampling interval) is determined for an adaptive control chart. When the process is in control, the sample size and sampling interval of an adaptive control chart should be the same as a traditional \bar{X} control chart. That is,

$$E[t(i)] = n_0 \quad (6)$$

$$E[t(i)] = t_0 \quad (7)$$

- (2) Design of parameters

- (i) Let I_1 be the zone within the warning limits, I_2 be the zone between the control limits and warning limits, and I_3 be the zone within the control limits. The lower and upper boundaries of I_1 , I_2 and I_3 are defined as follows:

$$I_1 = [-w, w] \quad (8)$$

$$I_2 = (LCL, -w) \cup (w, UCL) \quad (9)$$

$$I_3 = (LCL, UCL) \quad (10)$$

- (ii) From the assumption (ii), the following equation can be obtained:

$$n_0 \Pr(Z \in I_3) = n_1 \Pr(Z \in I_1) + n_2 \Pr(Z \in I_2) \quad (11)$$

Because Z follows a standard normal distribution, the following results can be obtained:

$$\Pr(Z \in I_3) = 2f(w) - 1$$

$$\Pr(Z \in I_2) = 2(f(UCL) - f(w)) \quad (12)$$

$$\Pr(Z \in I_1) = 2f(UCL) - 1$$

Using (11) and (12), the warning limits can be rewritten as follows:

$$w = f^{-1} \left[\frac{2f(UCL)[n_0 - n_2] + n_1 - n_0}{2(n_1 - n_2)} \right] \quad (13)$$

For a given sampling interval, a formula derived from (11) and (12) is given as follows:

$$t_0 \Pr(Z \in I_3) = t_1 \Pr(Z \in I_1) + t_2 \Pr(Z \in I_2) \quad (14)$$

w can then be rewritten as follows:

$$w = f^{-1} \left[\frac{2f(UCL)[t_0 - t_1] + t_2 - t_0}{2(t_2 - t_1)} \right] \quad (15)$$

After obtaining w , fix any three of the four parameters n_1, n_2, t_1 and t_2 , the last parameter can be calculated using (13) and (15). For example, fixing the values of n_1, n_2 , and t_1, t_2, b and c can be obtained by setting (13)=(15):

$$t_2 = \frac{t_0(n_1 - n_2) - b - t_1 c}{n_1 - n_2 - c},$$

$$b = \mathcal{A}(t_0 - t_1)(n_1 - n_2) f(UCL),$$

$$c = 2(n_0 - n_2) f(UCL) + (n_1 - n_0).$$

3. CONSTRUCT A HIERARCHICAL CONTROL PROCESS FOR CLUSTERED DEFECTS

This study utilizes the clustering analysis to adjust the number of defects on a wafer when the cluster phenomenon is detected. By doing so, the defect clusters are distributed randomly on a wafer and the type I error of the c-chart can then be reduced. The uniformity of a wafer is calculated and the adaptive control chart is employed to detect the wafer-to-wafer and lot-to-lot variations. The proposed procedure for detecting the clustered defect is described in the following:

Step 1: Obtain the wafer map using KLA 2110 wafer inspection system.

Cluster analysis is utilized to combine the clustered defects. Treat all defects in a cluster as one defect and the location of this defect is the location of the cluster's center. Recalculate the number of defects on each wafer

Use a goodness of fit test to check whether the adjusted (reduced) number of defects satisfies the assumption of a Poisson distribution. If it does, go to Step 2; otherwise, repeat Step 1.

Step 2: Determine whether the wafer-to-wafer variation is in control.

Calculate the value of NU using the adjusted number of defects on a wafer. Transfer the values of NU to satisfy the normality assumption. The $\bar{X}_{NU} - R_{NU}$ chart and $X - MR$ chart are constructed to detect the wafer-to-wafer variation.

Step3: Determine whether the lot-to-lot variation is in control.

The proposed adaptive U control chart is developed based on the \bar{X} control chart. In monitoring IC manufacturing process, the wafers are usually sampled from every lot. The sampling interval is treated as a fixed constant (e.g., a lot). Therefore, only the sample size needs to be changed according to the previous sampling result. The process of using U control chart to monitor the lot-to-lot variation is described as follows:

1. Standardize the sample data. Because m and s are unknown, they are estimated using the historical data. The transformation formula is :

$$Z_i = \frac{u_i - m}{\sqrt{s^2/n_i}} \cong \frac{u_i - \bar{u}}{\sqrt{\bar{u}/n_i}} \quad (16)$$

where \bar{u} is the mean number of defects on each wafer and the variance is the same as the mean since the number of adjusted defects on a wafer follows a Poisson distribution.

2. Contract different sampling schemes (strict and loose) for different sample size n_1 and n_2 .

3. Determine the control limits and the warning limits of the U control chart. The control limits are ± 3 since the sample data is standardized. The warning limits can be obtained using the following formula:

$$w = f^{-1} \left[\frac{2f(UCL)[n_0 - n_2] + n_1 - n_0}{2(n_1 - n_2)} \right] \quad (17)$$

4. Plot the sample points on the U control chart. If a sample point falls outside of the control limits, the process is said to be out of control. If a sample point falls between the control limits and the warning limits, n_2 is adopted as the next sample size. If a sample point falls within the warning limits, n_1 will be used as the next sample size.

5. Return to Step2 to perform the next sampling.

4. CASE STUDY

This section presents a case study to demonstrate the effectiveness of the proposed approach. The required data in this case study are obtained from an IC manufacturing company in Taiwan. When the wafer goes through the process of "Metal 2 Etch," the

coordinates of the wafer’s defects are collected using a KLA 2110 wafer inspection system which is a standard inspection tool utilizing the laser scan technique. Follow the steps described in Section 3, the result of each step is given as follows.

Step 1 : There are forty lots of wafers and each lot has twenty-five wafers. Two wafers, the fifth and the twenty-fifth wafers, of every lot were selected. The coordinates of defects and the number of defects are obtained for each wafer.

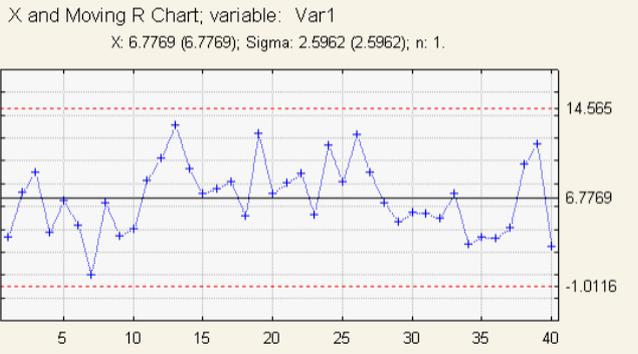


Fig. 1. The revised \bar{x} control chart

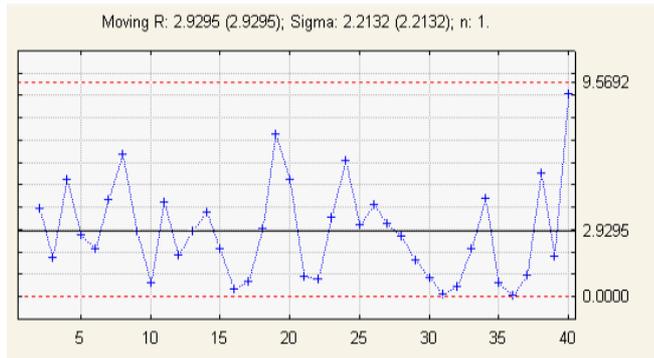


Fig.2. The revised MR control chart

Step2: The $\bar{X}_{Nu} - R_{Nu}$ chart and $X - MR$ chart are constructed to detect the wafer-to-wafer variation. The results are shown in Fig, 1 and 2. No wafer-to-wafer variations are observed in these two charts.

Step3 :

$$w = f^{-1} \left[\frac{2f(3)[1.5-2]+1-1.5}{2(1-2)} \right]$$

$$= f^{-1}[0.74935]$$

$$= 0.675$$

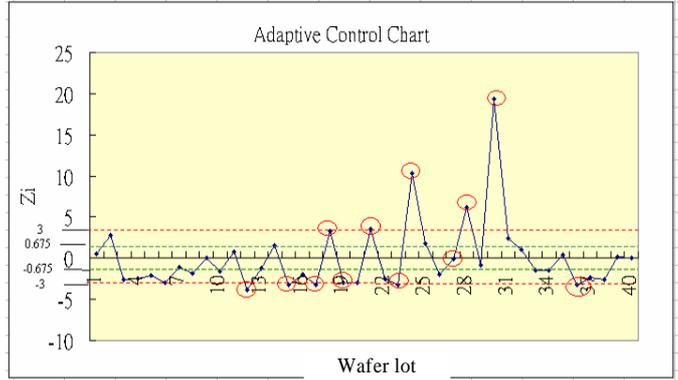


Fig 3. The proposed \bar{U} adaptive control chart

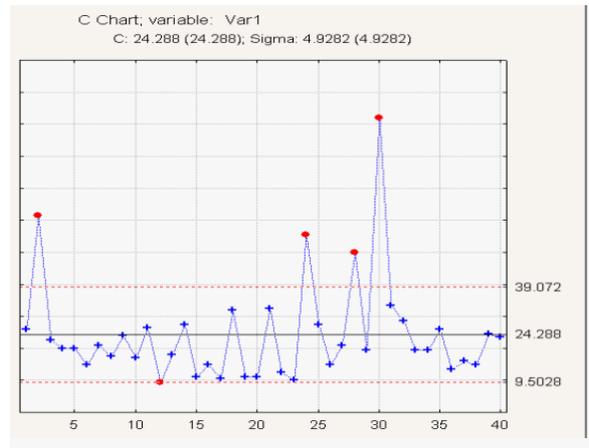


Fig.4. The traditional c-chart

The \bar{U} adaptive control chart can detect 11 out-of control lots; while the traditional c-chart only detect 5 out-of-control points. Consequently, the proposed control process can detect small shift and determine the source of variation coming from the wafer-to-wafer variation or the lot-to-lot variation.

5. CONCLUSION

As the wafer size increases, a wafer’s defects are no longer randomly distributed. They tend to cluster. In this study, a procedure is proposed to monitor the wafer defects in IC manufacturing. The Clustering Analysis is employed to adjust the number of wafer defects, thereby allowing not only the adjusted number of defects to satisfy the assumption of Poisson distribution but the conventional c-chart to still be used as well. The adaptive control chart is utilized to reduce the sampling cost. The proposed approach can reduce the false alarms caused by the clustered defects. The effectiveness of the proposed approach is demonstrated through a case study.

REFERENCES

[1] Costa, F. B. Antonio, “Joint \bar{X} and R Charts with Variable Sample Size and Sampling Intervals”, Journal

of Quality Technology , Vol.31 , No. 4, pp387-398

[2] Costa, F. B. Antonio, “ \bar{X} Charts with Variable Parameters”, Journal of Quality Technology ,Vol. 31 ,No.3, PP408-416, 1999

[3] Prabhu, S. Sharad, Montgomery, C. Douglas and Runger, C. George, “A Combined Adaptive Sample Size

and Sampling Intervals \bar{X} Control Scheme”, Journal of Quality Technology, Vol. 26 ,No.3, pp164-176, 1994

[4] Sharma, Subhash, “Applied Multivariate Techniques”, Wiley , 1996

[5] Wells, W. Stephen and Smith, D. James, “Making Control Chart Work For You”, Semiconductor

Decision Making and Analysis for Unexpected Road Blockages 1

Bing Su², Yinfeng Xu^{2,3}

²School of Management, Xi'an Jiaotong University, Xi'an 710049, China

³The State Key Lab for Manufacturing Systems Engineering, Xi'an 710049, China
subing@mailst.xjtu.edu.cn, yfxu@mail.xjtu.edu.cn

ABSTRACT

The unexpected road blockage problem (URBP for short) is considered for the case when some blockages occur on the road at certain times and such blockages would be revealed only upon reaching them. Papadimitriou and Yannakakis proved that devising a strategy that guarantees a given competitive ratio is PSPACE-complete if the number of edges that might be blocked is not fixed; Bar-noy and Schieber considered the deterministic and stochastic variations of the recoverable-URBP in the worst-case criterion. In this paper, we present an offline algorithm for the recoverable-URBP optimal solution and the complexity of the algorithm is $O(n^2)$ (n is the number of nodes). Two online strategies, waiting strategy and the Greedy strategy, are proposed and the competitive ratios of the two strategies are given. Furthermore, we compare the two strategies based on competitive ratio analysis.

Keywords: decision making, unexpected road blockage, offline algorithm, competitive ratio

1. INTRODUCTION

The unexpected road blockage problem (URBP for short) can be stated as follows, a traveler knows a graph (a map) $G(V, E)$, and the traveler has to go from s (origin) to t (destination). However, the map is unreliable, some blockages may occur on the road at certain times (blocked by unexpected events).

The unexpected road blockage problem is first introduced by Papadimitriou and Yannakakis [1], they proved that if the number of roads that might be blocked is not fixed, and then devising an online algorithm with a bounded competitive ratio is PSPACE-complete. Bar-Noy and Schieber [2] studied several variations of the problem. In the recoverable-URBP, each site is associated with a recovery time to reopen any blocked road that is adjacent to it, they presented a polynomial-time traveling strategy that guarantees the shortest worst-case travel time, in the case where an upper bound on the number of blockages is known in advance, and the recovery times are not long relative to the travel times. For the stochastic recoverable-URBP, they present an $O(m \log n)$ (n and m are the number of nodes and edges) time algorithm for designing a travel strategy from all sites to a fixed destination that guarantees the shortest expected travel time, under the assumption that the recovery time of the blocked edges are smaller than the edge's passing time. For the deterministic recoverable-URBP, they gave an $O(k^2 m + kn \log n)$ (k is the number of that road blockages) time algorithm for designing a travel strategy from all sites to a fixed destination guarantees the shortest worst-case travel time, under the assumption that the recovery time of the blocked edges are smaller than the edge's passing time. But, the assumption made in [2] is impractical for most transportation process. In the Irrecoverable URBP, once

a road is blocked, it remains blocked forever. Y.Xu et al [3] proposed two strategies, greedy strategy and reposition strategy, and showed that $2k + 1$ and $2^{k+1} - 1$ are their tight competitive ratios [4] [5] respectively, and proved that no online strategy has competitive ratio less than $2k + 1$ for Irrecoverable online URBP and in most cases, greedy strategy is better than reposition even though the competitive ratio of the former strategy is greater than the later ones.

In this paper, we focus on the recoverable URBP. We present an offline algorithm for the optimal solution based on analysis of the blocked edges cost (time) which is effected by the different starting time and end time of blockages, and the memory makes the problem be unfit for a converse solution of dynamic programming. The algorithm is a modification to the Dijkstra algorithm [6] and the complexity of the algorithm is $O(n^2)$. Then, according to the different assumptions on the recovery time, we present two different strategies, waiting strategy and the Greedy strategy; we show the competitive ratios of the two strategies. Furthermore, the comparison between these two strategies is discussed based on competitive ratio analysis.

2. PROBLEM STATEMENTS AND FORMULATION

Suppose the traveler has to go from s to t . He knows a map that consists of set of roads between two sites. The traveler will choose a shortest road (SR for short) from s to t . At certain times, the SR will be blocked by some events, such as traffic accidents. Assume that the blockages are recoverable. If all of the blockages' position and its recovery time are known in advance, the problem is considered as an offline

problem. If the blocked edges occur one by one without any predictable information, the problem is considered as an online problem.

3. AN OFFLINE ALGORITHM

Let $G(V, E)$ denote an edge weighted graph with $|V|$ vertices. v_1 denotes the origin and v_n the destination, $E = \{e_{ij} / (v_i, v_j), v_i, v_j \in V, i \neq j\}$ denote the set of the edges between the two sites, $T(v_1)$ denotes the starting time of v_1 , $T(e_{ij})$ denotes the passing time of e_{ij} without the blockage, $T'(e_{ij})$ denotes the recovery time segment of the blockage on e_{ij} , $T'(e_{ij}) = [t_i', t_i'']$, t_i' denotes starting time of the blockage, t_i'' denotes the end time of the blockage.

For convenience of the discussion, we do not distinguish the cost and the time for the traveler. In addition, all discussions are based on the following essential assumptions:

- (1) The traveler only stays at the vertices of G .
- (2) The blockage occurs on the e_{ij} only once.
- (3) $T_1(v_1) = 0$

3.1 The analysis of waiting time $T_1'(e_{ij})$

Let t_i denote the reaching time of the v_i who is the starting vertex of e_{ij} , $T_1'(e_{ij})$ denotes the waiting time of e_{ij} . The waiting time $T_1'(e_{ij})$ is different from the blockages' recovery time $T(e_{ij})$. we consider the several cases of $T_1'(e_{ij})$ under the different starting time and end time of blockages as follows :

- (1) If the starting time of the blockage occurs in $[0, t_i]$ ($t_i' < t_i'' < t_i$) then $T_1'(e_{ij}) = 0$
- (2) If the starting time of the blockage occurs in $[t_i, t_i'']$ ($t_i \leq t_i' \leq t_i''$), and $t_i' \leq T(e_{ij}) + t_i$ then $T_1'(e_{ij}) = t_i'' - t_i$
- (3) If the starting time of the blockage occurs in $(t_i, t_i'']$ ($t_i < t_i' \leq t_i''$), and $t_i' > T(e_{ij}) + t_i$ then $T_1'(e_{ij}) = 0$
- (4) If the starting time of the blockage occurs in $(t_i', t_i'']$ ($t_i' < t_i < t_i''$) then $T_1'(e_{ij}) = t_i'' - t_i$

From the above analysis, we can obtain a lemma as follows :

Lemma 1 If the blockages are recoverable, then the

passing time of the blocked edge e_{ij} is only relative with the t_i'' and $T(e_{ij})$, but irrelative with t_i' .

3.2 Algorithm analysis

Let $T_1(v_n)$ denote the shortest time from v_1 to v_n , $P(v_i)$ (the shortest time from v_1 to v_i) denote the permanence label of v_i , L (the upper bound for the passing time of the shortest paths from v_1 to v_i) denote the temperament label of v_i .

Step 1:

Let $S_0 = \{v_1\}$, $i = 1, \bar{S} = V \setminus S_i (= \bar{L}, \quad)$
 (\bar{S} denotes the adjacent vertices set of v_i)

and

$$\begin{cases} T_1(v_1) = 0 \\ L(v_j) = \infty \end{cases} \quad v_j \in \bar{S}$$

Step 2

If $v_n \in S_i$, then end the computation, otherwise, turn to the next step.

Step 3

If $v_j \notin S_i$ and $e_{kj} = (v_k, v_j) \in E$, from the analysis in section 3.1, compute the $T_1'(e_{kj})$ and $T(e_{kj}) + T_1'(e_{kj})$

Step 4

If $L(v_j) \geq P(v_k) + (T(e_{kj}) + T_1'(e_{kj}))$, then modify the $L(v_j)$ to $P(v_k) + (T(e_{kj}) + T_1'(e_{kj}))$, Otherwise, turn to the next step.

Step 5

Let $L(v_{i_i}) = \min_{v_j \in S} \{L(v_j)\}$,

if $L(v_{i_i}) \leq +\infty$, then update the L of v_{i_i} to P ,

and let $P(v_{j_i}) = L(v_{j_i})$ and

$S_{i+1} = S_i \cup \{v_{j_i}\}$ ($k = j_i$), then, update i to $i + 1$ and turn to step 2, otherwise, end the computation. At this time, if $v \in S_i$ then $T_1(v) = P(v)$, if $v \notin S_i$ then $T_1(v) = L(v)$.

3.3 The complexity of algorithm analysis

Let n and m be the number of vertices and edges

in $G(V, E)$. In the worst-case criterion, when the traveler reaches the vertex v_i , and find all of edge which starting vertex is v_i have the blockages, we know that the computation times of $T_1'(e_{ij})$ is m , the computation times of step3 is $2m$ and the computation times of step3 is $O(n^2)$ at most. From the above analysis, we know that the complexity of the algorithm is $O(n^2)$.

Theorem 1 The algorithm can compute the passing time of shortest path from v_1 to v_n within $O(n^2)$ time.

4. THE ONLINGE STRATEGY

Since the blockages are recoverable, when the traveler reaches the blockage, he may choose such strategies: one is that he stays at the blockage and waits for reopening of the road, then go on along the SR . It is reasonable for the traveler to wait until the road is reopened if the recoverable time is not too long. Another strategy that he chooses an efficient path goes along a new path or waits until the road reopening.

Let $G(V, E)$ denote an edge weighted graph with $|v|$ vertices, s denote the origin and t denote the destination, $\delta(e_1, e_2, \dots, e_k)$ $e_i = (x_i, y_i)$ ($i = 1, 2, \dots, k$) denote the blocked edges sequences, x_i ($i = 1, 2, \dots, k$) denote the beginning vertex of e_i , y_i ($i = 1, 2, \dots, k$) denote the ending vertex of e_i , $t'(e_i)$ denote the recovery time for the blockage happened at e_i , $t(e_i)$ denote the passing time of the e_i , $S(s, t)$ denote the cost (time) from s to t without any blockage, $x_i - t$ denote the road from x_i to t on SR .

For convenience of the discussion, we do not distinguish the cost and the time for the traveler. In addition, all discussions are based on the following essential assumptions:

- (1) G is connected even when the blocked edges are removed.
- (2) The blockage information is available to the traveler only when he reaches it, and the passed path cannot be blocked.
- (3) When the traveler meets a blockage, he may obtain the information about the recovery time.

For the classical online problem, the competitive ratio is a constant irrelative with the sequences of events. Let

$C_{opt}(R)$ denote the cost of the optimal offline problem, $C_A(R)$ denote the cost of the strategy A for the online problem. Strategy A is called α -competitive^{[7][8]} if the following inequality holds $C_A(R) \leq \alpha \cdot C_{opt}(R) + \beta$, where α and β are constants irrelative with R . Papadimitriou and Yannakakis proved that if the number of roads that might be blocked is not fixed, and then the online algorithm with a constant competitive ratio does not exist. For the URBP, we give a new definition for the function competitive ratio. Let $C_{opt}(\delta)$ denote the cost of the optimal solution for the offline case for the traveler to go from s to t , $C_A(\delta)$ denote the cost of the online strategy A for the traveler to go from s to t .

Strategy A is called $f(k)$ -competitive, if the following inequality holds

$$C_A(\delta) \leq f(k) C_{opt}(\delta)$$

where $f(k)$ is relative with k .

We will analyze the performance of the different strategies and show the competitive ratio in the following sections.

We use the following reasonable fact to bind the optimal solution.

FACT: $C_{opt}(\delta)$ is no less than $S(s, t)$.

4.1 The Waiting Strategy

WS (Waiting Strategy): when the traveler reaches e_i (along the SR) and finds that e_i is blocked and cannot pass through within $t(e_i)$, he choose to stay at x_i and wait for reopening of the road, then go on along the x_i . Denote this strategy as W .

The total cost of the traveler from s to t with W is

$$C_W(\delta) = S(s, t) + \sum_{i=1}^k t'(e_i) \tag{4.1}$$

From (4.1) the following equality holds

$$C_W(\delta) = \left[1 + \frac{\sum_{i=1}^k t'(e_i)}{S(s) t} \right] S(s) t \tag{4.2}$$

In order to get clearer, we make the following assumptions and all the assumptions are in accordance with the reality.

Assumption 1 $t'(e_i) \leq \alpha \cdot t(e_i)$ for any e_i .

($\alpha_i, i = 1, 2, \dots, k$) denotes the ratio of $t'(e_i)$ to $t(e_i)$, $\alpha = \max\{\alpha_1, \alpha_2, \dots, \alpha_k\}$

As

$$\sum_{i=1}^k t(e_i) \leq S(s, t)$$

we have

$$\sum_{i=1}^k t'(e_i) \leq \alpha S(s, t)$$

Substitute into (4.1), the following inequality holds:

$$C_W(\delta) \leq (1 + \alpha) S(s, t) \quad (4.3)$$

From the above analysis, we can obtain the following theorem:

Theorem 2 If $t'(e_i) \leq \alpha \cdot t(e_i)$ holds, then the competitive ratio of W is $1 + \alpha$.

According to the definition of the competitive strategy and competitive ratio, we could conclude that W is a competitive strategy with competitive ratio $(1 + \alpha)$ for online recoverable URBP.

Similar, if $\alpha = 1$, from theorem 2, we have

Corollary 1 If $t'(e_i) \leq t(e_i)$ holds, then the competitive ratio of W is 2.

From the above discussion, it can be found that the competitive ratio of Waiting strategy only related to α (the ratio between recovery time $t'(e_i)$ and passing time $t(e_i)$), and it has no relation with the blockages numbers and their positions. When $\alpha = 1$, the travel time would not exceed the two times of the $S(s, t)$.

For waiting strategy, let $S(x_i, t)$ denote the passing time from x_i to t when the i -th blocked edge occurs on $S(s, t)$. Because the blocked edges always happen at the SR , so we could further relax our assumptions.

Assumption 2 $t'(e_i) \leq \alpha \cdot S(x_i, t)$ for any e_i .

($\alpha_i, i = 1, 2, \dots, k$) denotes the ratio of $t'(e_i)$ to $S(x_i, t)$, $\alpha = \max\{\alpha_1, \alpha_2, \dots, \alpha_k\}$

For

$$S(x_i, t) \leq S(s, t)$$

then

$$\sum_{i=1}^k t'(e_i) \leq k \alpha S(s, t)$$

Substitute into (4.1), the following inequality holds:

$$C_W(\delta) \leq (1 + k \alpha) S(s, t)$$

Theorem 3 If $t'(e_i) \leq \alpha \cdot S(x_i, t)$ holds, then the competitive ratio of W is $1 + k \alpha$

Similarly, if $\alpha = 1$, from theorem 1, then, we have

Corollary 2 If $t'(e_i) \leq S(x_i, t)$ holds, then the competitive ratio of W is $k + 1$.

From the proof of theorem 3, we would find that when the assumption is further relaxed, if

$t'(e_i) \leq \alpha \cdot S(x_i, t)$ then the theorem 3 and corollary 2 always hold.

4.2 The Greedy Strategy

GS(Greedy Strategy): when the traveler reaches e_i (along the SR) and finds that e_i is blocked and can not pass through within $t(e_i)$, he choose an efficient path, staying at x_i and waiting for reopening of the road and go on along the $x_i - t$ or choose a new shortest path from x_i to t without the blockages, whichever is better. Denote this strategy as G .

Let $L'(x_i, t)$ denote the cost of the road that the traveler hasn't finished. If the traveler waits at the x_i and goes through when the e_i is recovered, then the passing time would be denoted as $L'(x_i, t) + t'(e_i)$. If the traveler always finds a new shortest path from x_i to t without the blockages and denotes it as $L''(x_i, t)$, then the greedy strategy should choose the path with

$$\min\{L'(x_i, t) + t'(e_i), L''(x_i, t)\}$$

Denote $L(s, x_i)$ as the travel time before arriving e_i , $S(v_1, x_1)$ as the travel time from v_1 to x_1 on SR , $S(x_1, v_n)$ as the travel time from x_1 to v_n on SR , $C_G(\delta)$ as the total travel time from s to t by taking the greedy strategy. For simplicity, we suppose that every time the traveler chooses a new optimal path without waiting when he encounters the blocked edges, we can obtain:

$$L''(x_i, t) \leq L'(x_i, t) + t'(e_i) \quad (i = 1, 2, \dots, k)$$

In the worst case, all the blocked edges happen at the chosen optimal path and we can obtain:

$$C_G(\delta) = L(s, x_i) + L''(x_i, t) \quad (i = 1, 2, \dots, k)$$

Assumption 1 $t'(e_i) \leq \alpha \cdot t(e_i)$ for any

$e_i, (\alpha_i, i = 1, 2, \dots, k)$ denotes the ratio of $t'(e_i)$ to $t(e_i)$, $\alpha = \max\{\alpha_1, \alpha_2, \dots, \alpha_k\}$

If there is only one blocked edge, we have

$$C_G(e_1) = L(s, x_1) + L''(x_1, t)$$

Since

$$L''(x_1, t) \leq L'(x_1, t) + t'(e_1)$$

$$t'(e_1) \leq \alpha \cdot t(e_1)$$

and

$$t(e_1) \leq L'(x_1, t)$$

$$L(s, x_1) = S(s, x_1)$$

$$L'(x_1, t) = S(x_1, t)$$

We obtain

$$L''(x_1, t) \leq (1 + \alpha)S(x_1, t)$$

$$\begin{aligned} C_G(e_1) &= S(s, x_1) + (1 + \alpha)S(x_1, t) \\ &= S(s, t) + \alpha S(x_1, t) \end{aligned}$$

$$S(x_1, t) \leq S(s, t)$$

$$C_G(e_1) \leq (1 + \alpha)S(s, t)$$

When encountering the second blocked edge, we also have

$$\begin{aligned} C_G(e_1, e_2) &= L(s, x_2) + L''(x_2, t) \\ &\leq L(s, x_2) + L'(x_2, t) + t'(e_2) \end{aligned}$$

Since

$$L'(x_2, t) = L''(x_1, t) - L''(x_1, x_2)$$

$$L(s, x_2) = L(s, x_1) + L''(x_1, x_2)$$

we have

$$C_G(e_1, e_2) \leq L(s, x_1) + L''(x_1, t) + t'(e_2)$$

Since $t'(e_2) \leq \alpha \cdot t(e_2)$

$$t(e_2) \leq L''(x_1, t)$$

we have

$$C_G(e_1, e_2) \leq L(s, x_1) + (1 + \alpha)L''(x_1, t)$$

Since

$$L''(x_1, t) \leq (1 + \alpha)L'(x_1, t)$$

$$L(s, x_1) = S(s, x_1)$$

$$L'(x_1, t) = S(x_1, t)$$

we have

$$C_G(e_1, e_2) \leq (1 + \alpha)S(s, t)$$

Similarly, it could be concluded that

$$C_G(e_1, e_2, \dots, e_k) \leq (1 + \alpha)S(s, t)$$

Therefore the following theorem holds :

Theorem 4 If $t'(e_i) \leq \alpha \cdot t(e_i)$ holds, then the competitive ratio of G is $(1 + \alpha)^k$.

Similar, if $\alpha = 1$, from Theorem 4, we have

Corollary 3 If $t'(e_i) \leq t(e_i)$ holds, then the competitive ratio of G is 2^k .

From the proof of theorem 4, we would find that when the assumption is further relaxed, if

$$t'(e_i) \leq \alpha \cdot S'(x_i, t), t'(e_i) \leq \alpha \cdot S'(s, t)$$

and corollary 3 always hold.

4.3 Comparisons between two strategies

Table 1 Competitive ratio between Two Strategies

Recovery time assumption	Competitive ratio	
	Waiting strategy	Greedy strategy
$t'(e_i) \leq t(e_i)$	2	2^k
$t'(e_i) \leq \alpha \cdot t(e_i)$	$1 + \alpha$	$(1 + \alpha)^k$
$t'(e_i) \leq S(s, t)$	$k + 1$	2^k
$t'(e_i) \leq \alpha \cdot S(s, t)$	$1 + k \cdot \alpha$	$(1 + \alpha)^k$

From the above analysis, within the assumption $t'(e_i) \leq \alpha \cdot t(e_i)$, we know that the competitive ratio of W does not relate itself to the number of the blockages,

if $0 < \alpha < 1$, then $S(s, t) \leq C_W(s, t)$. In this case, the total

waiting (cost) time is very little comparing with the moving cost on SR (just as a red light for the traffic control), this means the W strategy is more efficient and the total cost (time) taken by using strategy W approaches the cost (time) taken by an offline optimal.

If $\alpha > 1$ then $C_W(s, t) \leq (1 + \alpha)S(s, t)$. In this case, the total (cost) time of waiting may be very long (such as a traffic accident happens), but it needs to consider in more details to evaluate the performance of the W strategy.

From (4.3), we see that the competitive ratio is mainly dependent on the ration between the total waiting time and the $S(s, t)$, not only on α , this means that in some cases, even if α is very larger, but the performance of W may still be very efficient. For the worst case (all α_i are the same and α is a large), W is not as efficient as the traveler expected.

We know that the competitive ratio of greedy strategy increases exponentially in anyone assumption, so, even in the worst case, waiting strategy is better than greedy strategy when the blocked edges always happen on the newly chosen path. However, people always would like to choose the greedy strategy when encounter the jam. They think that the greedy strategy is efficient and the cost of moving is less than waiting. In fact, this choice reflects that people do not have the correct expectation on unexpected crisis. Since greedy is widely acquired, so in what situation is greedy strategy superior to

waiting strategy would be the future direction of the research.

5. CONCLUSION

As discussed in the paper, we present an offline algorithm for the Recoverable URBP based on that the blocked edges cost (time) is effected by the different starting time and end time of blockages, and the memory makes the problem be unfit for a converse solution of dynamic programming. The algorithm is a modification to the Dijkstra algorithm, and the complexity of the algorithm is $O(n^2)$. The optimization theory always treats the problem from perspectives of stand-by and assumes that the constraints do not vary, which is always non-realistic. The competitive algorithm may shed some light on the optimization of the problem taking the variations of critical factors. We study the online strategies when the blockages that occur one by one without any predictable information except whose recovery time are known. By making different assumptions on the recovery time of blocked edges, two strategies, waiting strategy and the greedy strategy are proposed and the competitive ratios of the two strategies are given. Furthermore, by comparing these two strategies, it could be found that waiting strategy is superior to greedy strategy in the same assumption. For the recoverable URBP, there have some further direction to work, such as how to deal with the problem when blocked happen at the vertices.

ACKNOWLEDGEMENT

The authors would like to acknowledge the support of research grant No.10371094, 70121001 from the NSF of China.

REFERENCES

- [1]C.H.Papadimitriou and M.Yannakakis, "Shortest paths without a map", *In Proc. 16th ICALP, Lect. Notes in Comp. Sci.*, No.372, pp.610-620,1989.
- [2]A.Bar-Noy, B.Schieber, "The Canadian Traveler Problem", *Proceedings of the second annual ACM-SIAM symposium on Discrete algorithms*, pp.261-270,1991.
- [3]Y.Xu, Z.Zhu and B.Su. "Online Canadian Traveller Problem and its Competitive Analysis", submitted,2004.
- [4]M.S.Manasse, L.A.McGeoch, and D.D.Sleator, "Competitive algorithms for server problems", *Journal of Algorithms*, Vol.11,208-230,1990.
- [5]S.Ben David and A.Borodin, "A new measure for the study of the on-line algorithm", *Algorithmica*, Vol.11, 73-91. 1994.
- [6]E.W. Dijkstra. "A note on two problems in connexion with graphs Nume" *r. Math* 1959. pp. 269-271.
- [7]D.S.Hochbaum, "Approximation Algorithms for NP-Hard Problems", *PWS publishing company*, 1997.
- [8]Y.F.Xu, K.Wang, J.Ding, "Online k-taxi problem and competitive analysis on restricted graph", *Journal of System Engineering*, Vol14, NO.4, pp.76-82,1999.

Discovering Fuzzy Functional Dependencies as Semantic Knowledge in Large Databases

Xue Wang , Guoqing Chen

School of Economics and Management, Tsinghua University, Beijing 100084, China
 {wangx3.03, chengq}@em.tsinghua.edu.cn

ABSTRACT

Fuzzy functional dependency (FFD) is a kind of semantic knowledge and can be discovered from a large volume of business data. Sectional FFD and Attribute FFD are discussed so as to reflect semantics of the business world and express useful information that is natural for people to comprehend. The experimental results on an insurance data set show that the proposed method can extract knowledge efficiently and effectively.

Keywords: data mining, knowledge discovery, fuzzy functional dependency, Sectional FFD, Attribute FFD.

1. INTRODUCTION

Widespread internet applications and e-business practices have resulted in a rapid growth of the database size that is beyond the scope of expert human capabilities to scan all the collected data and to discover the useful knowledge hidden in it. Therefore, data mining and knowledge discovery, for finding interesting patterns, dependencies, summaries, regularities, etc, has become an increasingly important subject in the research area of database and business intelligence.

of the data. For example, we may get a FFD at the attribute level (Attribute FFD), such as Age ~> Salary (Age fuzzy determines Salary), but some times, we may only get FFD for sub-classes of attribute values (Sectional FFD), such as Age(young) ~> Salary(low). Hence the construction and discovery of Sectional FFD is useful and novel and can serve as a type of interesting pattern and aid to enrich the knowledge base of a company.

2. SECTIONAL FFDs AND ATTRIBUTE FFDs

Functional dependency (FD) is important in both database design and maintenance. The classical definition of FD is: X functionally determines Y, (or Y is functionally dependent on X), denoted by X→Y, if and only if $\forall t_1, t_2 \in R$, if $t_1(X) = t_2(X)$ then $t_1(Y) = t_2(Y)$. But in real world applications, information is often incomplete or ambiguous. For example, customers may not be willing to provide their actual age but "about twenty", "middle aged", "25-30" or the like. With the inception of fuzzy logic [13] to model imprecise information, fuzzy extensions has been introduced into functional dependency and different aspects since 1980's (Prade et al. [5] Raju et al.[6] Bhuinya et al.[1], Chen et al.[2][3], Cubero et al[4], Saxena et al.[7], S.Ben Yahia et al.[11], etc). A general setting of fuzzy functional dependencies proposed by Chen et al is as follows [2]:

In this section, we discuss the notions and properties of Sectional FFDs and Attribute FFDs, viewed as a kind of semantic knowledge.

2.1 Notions

When dealing with fuzzy data, linguistic labels are defined in a unified way with membership grades. Examples of these labels are small, large and young. In some situations, such labels are used to represent abstract and linguistic summarization of quantitative data values. Thus, the data concerned can be represented in the form of a vector like $V(\mu_1/L_1, \mu_2/L_2 \dots \mu_n/L_n)$, where $L_1 \dots L_n$ are labels and $\mu_1 \dots \mu_n$ are membership grades. Next, we replace this vector with the maximal membership grade and its corresponding label μ_i where $\mu = \max(\mu_1 \dots \mu_n)$. For instance, "about 50" can be translated into "0.9/old".

Let U be the set of all attributes for a relation scheme R and $X, Y \subseteq U$. X functionally determines Y (or Y is functionally dependent on X) to the degree ϕ , denoted by $X \rightarrow_{\phi} Y$, if and only if for any tuples t_1, t_2 ,

$$\min_{t_1, t_2 \in R} |c(X(t_1), X(t_2)), c(Y(t_1), Y(t_2))| \geq \phi \quad (1)$$

where $\phi \in [0,1]$ I is a fuzzy implication operator (FIO) and c is an equality measure.

Definition 1 Let U be the set of all attributes for a relation scheme R, and $X, Y \subseteq U$, $L_1 \dots L_n$ are labels of X and $L'_1 \dots L'_n$ are labels of Y. For a tuple t in R, its X and Y values are μ_i/L_i and μ'_j/L'_j . $\theta, \alpha \in [0,1]$, where α is a given threshold. t satisfies a tuple relation to the degree θ , denoted by $X(t, \mu_i) \sim_{\theta} Y(t, L'_j)$ if and only if

$$\theta = I(\mu_i, \mu'_j) \geq \alpha \quad (2)$$

In a similar spirit, our previous work [12] presented a specific type of fuzzy functional dependency based on tuples and label closeness measures, which has a good arithmetic efficiency in the context of data mining. In addition, it can reflect both overall and partial knowledge

where I is a fuzzy implication operator (FIO). Most used FIOs are Lukasiewicz operator, Kleene-Dienes operator, Gödel operator, R operator, etc. Here, we choose R

implication operator [8] as an example to derive a specific form of tuple relation which has some good properties as stated in section 2.3.

$$I_{R_0}(a, b) = \begin{cases} 1 & \text{if } a \leq b \\ \max(1-a, b) & \text{otherwise} \end{cases} \quad (3)$$

Definition 2 Label closeness measure which describes the relationship between two given labels denoted by $\sigma(L_i, L_j)$, satisfies the following properties:

- 1) $\sigma(L_i, L_i) = 1$ (4)
- 2) $\sigma(L_i, L_j) = \sigma(L_j, L_i)$ (5)

More specifically, for $L^* = \{L_1 \dots L_n\}$, let $\Theta(L^*) = \min(\sigma(L_i, L_j) \mid L_i, L_j \in L^*)$ and for $L^{**} = \{L_1^*, L_2^* \dots L_n^*\}$, which is a set of sets $\Theta(L^{**}) = \min(\sigma(L_1^*), \sigma(L_2^*) \dots \sigma(L_n^*))$.

Definition 3 For any label L_i of X , all tuples satisfying the tuple relation $X(t, L_i) \sim_{\geq \theta} Y(t, L'_j)$ compose a set T_i . more formally, $T_i = \{t \mid X(t, L_i) \sim_{\geq \theta} Y(t, L'_j)\}$. All those T_i compose a set named LA Sectional FFD, denoted by $X(L_i) \sim_{\geq \theta} Y(L_i^*)$ is valid if and only if

$$\Theta(L_i^*) \geq \beta \quad (6)$$

where β is a given threshold between 0 and 1, $\theta = \min(\Theta)$.

Definition 4 An Attribute FFD (AFFD) $X \sim_{\geq \theta} Y$ holds if and only if for any two label L_i and L_j , the Sectional FFDs $X(L_i) \sim_{\geq \theta} Y(L_i^*)$ hold and

$$I'(\sigma(L_i, L_j), \sigma(L_i^*, L_j^*)) \geq \omega \quad (7)$$

where I' is a FIO and ω is a given threshold. $\Theta' = \min(\Theta)$.

Here we also use I' to derive a specific form of FFD. We can easily see, if L^* is a single element set, the Sectional FFD $X(L_i) \sim_{\geq \theta} Y(L_i^*)$ mentioned above is surely valid.

2.2 An Example

The following table is part of a fuzzy relational database. The membership functions of attribute labels are given in figure 2. (with $\omega=0.7, \alpha=0.7$)

T#	Age	Salary
t1	Young	1200
t2	35	0.8/low
t3	About 30	About 1500
t4	50-55	3800
t5	0.9/old	5000
t6	60	0.9/high

Table 1 Example data table

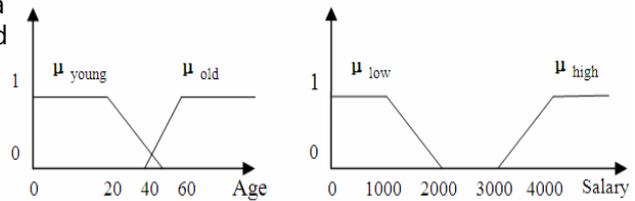


Figure 2 Membership functions

The first step: transform all values in table 1 into table 2 using maximal membership grade and its corresponding label.

T#	Age	Salary
t1	1/young	0.8/low
t2	0.75/young	0.8/low
t3	0.7/young	0.72/low
t4	0.75/old	0.8/high
T5	0.9/old	1/high
T6	1/old	0.9/high

Table 2 Transformed data

The second step: check if for every tuple, the tuple relation holds. We can easily derive that $Age(t_1, young) \sim_{\geq 0.8} Salary(t_1, low)$, $Age(t_2, young) \sim_{\geq 0.72} Salary(t_2, low)$, $Age(t_3, young) \sim_{\geq 0.72} Salary(t_3, low)$, $Age(t_4, old) \sim_{\geq 1} Salary(t_4, high)$, $Age(t_5, old) \sim_{\geq 0.9} Salary(t_5, high)$, $Age(t_6, old) \sim_{\geq 0.9} Salary(t_6, high)$. If the threshold $\alpha = 0.7$ then all tuple relations hold.

The third step: for attribute Age, label young has three tuples t_1, t_2 and t_3 , the corresponding label in attribute Salary $L^* = \{low\}$, L^* is a single element set, so the sectional FFD $Age(young) \sim_{\geq 0.72} Salary(low)$ holds. In the same way, we can easily derive $Age(old) \sim_{\geq 0.9} Salary(high)$, in addition, $\sigma(\sigma(young, old), \sigma(low, high)) = 1 > \omega$, then the Attribute FFD $Age \sim_{\geq 0.72} Salary$ holds.

Notably, if the salary of the first tuple is 0.8/high instead of 0.8/low, we will get $Age(t_1, young) \sim_{\geq 0.8} Salary(t_1, high)$. Therefore, for the left side label "young", $L^* = \{low, high\}$, we have to compute $\sigma(low, high)$. If $\sigma(low, high)$ is below the given threshold, the Sectional FFD $Age(young) \sim_{\geq \alpha} Salary(low, high)$ is denied.

2.3 Properties

It can be proved that Attribute fuzzy functional dependencies with R implication operator have the following important properties.

- 1) **Reflexivity:** if $Y \subseteq X \subseteq U$, where X is a set of attributes, then $X \sim_{\geq \alpha} Y$ holds. It is a trivial fuzzy functional dependency.
- 2) **Augmentation:** if $X \sim_{\geq \alpha} Y$ holds, then for $Z \subseteq U$, $XZ \sim_{\geq \alpha} YZ$ holds.
- 3) **Union Rule:** if $X \sim_{\geq \alpha} Y$, $X \sim_{\geq \beta} Z$, then $X \sim_{\geq \gamma} YZ$.
- 4) **Decomposition Rule:** if $X \sim_{\geq \alpha} Y$, $Z \subseteq Y$, where Y is a set of attributes, then $X \sim_{\geq \alpha} Z$.
- 5) **Partial Transitivity:** if $X \sim_{\geq \tau} Y$ and $Y \sim_{\geq \eta} Z$ with $\tau, \eta \geq 1/2$, then $X \sim_{\geq \gamma} Z$ holds with $\gamma = \min(\tau, \eta)$.

3. MINING ALGORITHM

In this section, we'll present a mining algorithm to discover Sectional FFDs and Attribute FFDs.

As shown in section 2.3, the AFFD $X \rightsquigarrow YZ$ can be decomposed in to $X \rightsquigarrow Y$ and $X \rightsquigarrow Z$. Therefore we will deliberate to find non-trivial AFFDs, each with a single attribute in its right-hand side. We can derive the minimal FFD set using a map listed below. For example, the line between X and XY means we will check whether the AFFD $X \rightsquigarrow Y$ holds, in the same way, the line between XY and XYZ means we will check whether the AFFD $XY \rightsquigarrow Z$ holds. An algorithm of the mining procedure for $X \rightsquigarrow Y$ is as follows.

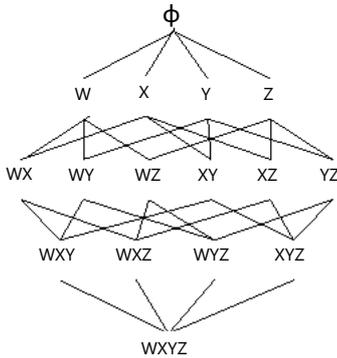


Figure 3 Map for mining

```

INITIALIZING:
for each Label  $l_k$  of X
     $L_k^* := \phi$ ;
    sectional_flag(k) := valid;
    attribute_flag := valid;
end for

MINING
for each t in R
    get corresponding X,Y value,  $\mu_i, \mu_j / L_j$ ;
    compute  $\theta := I(\mu_i, \mu_j)$ ;
    if  $\theta < \alpha$ 
        tuple relation is denied;
        sectional_flag(i) := denied;
    else
        add  $L_j$  into the set  $L_k^*$ ;
    end if
end for
for every label  $l_i$  of X
    if sectional_flag(i) = valid &  $\Theta(L_i^*) \geq \beta$ 
        the sectional FFD  $X(l_i) \rightsquigarrow Y(L_i^*)$  holds;
    else
        attribute_flag := denied;
    end if
end for
if attribute_flag = valid
    for every two labels  $l_i, l_j$  of X
        if  $I(\sigma(L_i, l_j), \sigma(L_i^*, L_j^*)) < \omega$ 
            attribute_flag := denied;
            break;

```

```

end if
end for
if attribute_flag = valid
    the Attribute FFD  $X \rightsquigarrow Y$  holds
end if

```

When mining such FFDs from databases, we need to check every tuple to see if it satisfies tuple relation and compute label closeness value. For $X \rightsquigarrow Y$, if X has M_1 labels and Y has M_2 labels, the algorithm's complexity is at $O(N + M_1^2 + M_2^2)$. In general, $M, M_1 < N$, so valuable patterns could be discovered quite efficiently.

4. EXPERIMENTAL RESULTS

We have carried out an experiment on a real business data set <http://www.smr.nl>. The algorithm was developed in C language, and run on PC with PIII866, RAM 256M, and Windows 2000 professional.

4.1 Data Description

The data set is a real set of an insurance company, provided by "Dutch Data Mining company sentient machine research". It contains 5822 tuples (transactions) and 86 attributes such as "number of houses", "customer main type", "average size household", etc. Because of space limitation, we only present the results discovered between the first 10 attributes within all tuples. Table 3 gives a general description of these first 10 attributes.

No	Description	Domain
1	Customer Subtype	1 High Income, expensive child ... 15 Household with children ... 41 Mixed rurals
2	Number of houses	Number of 1-10
3	Avg size household	1-6 (1 is the smallest, 6 the largest)
4	Avg age	1 20-30 years ... 6 70-80 years
5	Customer main type	1 Successful hedonists ... 8 Family with grown ups ... 10 Farmers
6	Roman catholic	0 0% 1 1 - 10% ... 9 100%
7	Protestant	0 0% ... 9 100%
8	Other religion	0 0% ... 9 100%
9	No religion	0 0%... ... 9 100%
10	Married	0 0% ... 9 100%

Table 3 Data Description

4.2 Data Preparation

Among the descriptions above, we can see that there are

fuzzy values, such as "Avg size household", graded knowledge is deemed interesting and novel, which may values, such as "Avg age" and crisp values, such as provide decision-makers with a better understanding of "Customer Subtype". We first transformed all fuzzy the customers. terms or graded figures into a unified type – linguistic labels with corresponding membership grades. For example, there are six ranks in the attribute of "Avg size household" – "1, the smallest and 6 the largest". WeIn this paper, we have discussed Attribute FFD and transfer them into 3 labels as "small, middle, large", and Sectional FFD as a kind of semantic knowledge which the original graded values can be translated as shown in an reflect overall or partial knowledge of the data in a manner that is natural for people to comprehend. The experimental results on an insurance data set showed that the proposed method can extract knowledge efficiently and effectively.

5. CONCLUSION

6. ACKNOWLEDGEMENT

The work was partially supported by the National Science Foundation of China (70231010, 70321001) and MOE Funds for Doctoral Programs (20020003095).

Rank	Label/value
1	small/1
2	small/0.9
3	medium/0.9
4	medium/1
5	large/0.9
6	large/1

Table 4 Fuzzy values of "Avg size household"

At the same time, we defined label closeness measures. For example σ (small, medium) = 0.8, σ (medium, large) = 0.8, σ (small, large) = 0.

4.3 Results

For illustrative purposes, we only list several sectional FFDs (Table 5) and Attribute FFDs (Table 6) with top θ values.

No	Sectional FFDs
1	Sub type [15] $\sim \rightarrow$ number of houses [many]
2	Sub type [15] $\sim \rightarrow$ age [old]
3	Sub type [15] $\sim \rightarrow$ protestant [medium]
4	age [young] $\sim \rightarrow$ number of houses [small]
5	Number of houses[huge] \rightarrow size of household [small]
6	Protestant [large] \rightarrow roman catholic [very small, small]

Table 5. Sectional FFDs

No	Sectional FFDs
1	sub type \rightarrow main type
2	number of houses \rightarrow size of household
3	size of household \rightarrow number of houses
4	size of household \rightarrow age
5	age \rightarrow number of houses
6	age \rightarrow size of household
7	main type \rightarrow number of houses
8	roman catholic \rightarrow number of houses
9	married \rightarrow number of houses

Table 6. Attribute FFDs

From the Attribute FFDs discovered, we can find some interesting patterns. For example, "number of houses" are determined by "size of household", "avg age", "main type", "roman catholic", "married", etc. Therefore "number of houses" is dependent on other attributes to a great extent. As to Sectional FFDs, they can reflect dependencies that hold partially. We can find that among Sectional FFDs, customers of "sub type 15" have many other behaviors in common. They are all old, have many houses, more than 50% of them are protestant. Such

REFERENCES

- [1] B. Bhuniya, P. Niyogi, Lossless join property in fuzzy relational databases, Data and Knowledge Engineering, 1993, pp. 109-124.
- [2] G.Q.Chen, J.Vandenbulcke, E.E.Kerre, A step towards the theory of fuzzy relational database design. Proceedings on Computer, Management science. IFSA'91, pp 44-47.
- [3] G. Q. Chen, E. E. Kerre, J. Vandenbulcke, Normalization based on fuzzy functional dependency in a fuzzy relationnal data model. Information Systems, Vol. 21, No.3, 299-310, 1996.
- [4] J.C.Cubero, M.A.Vila, A new definition of fuzzy functional dependency in fuzzy relational databases. International Journal of Intelligent Systems 9(5)(1994), 441-44.
- [5] H. Prade, C. Testemale, Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries, Information Sciences 34 (1984) 115-134.
- [6] K.S.V.N.Raju, A.K.Majumadar, Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database systems, ACM Transactions on Database Systems 13(2), 1988, 129-166
- [7] P.C. Saxena, B.K. Tyagi, Fuzzy functional dependencies and independencies in extended fuzzy relational databases models, Fuzzy Sets and Systems 69 (1995) 65-89.
- [8] Guojun Wang, A formal deductive system for fuzzy propositional calculus. Chinese science bulletin Vol.42 No.18, September 1997
- [9] S.L.Wang, J.W.shen, T.P.Hong, Mining fuzzy functional dependencies from quantitative data. 2000 IEEE
- [10] S.B.Yahia, A.Jaoua. Mining linguistic summaries of databases using based Lukasiewicz implication fuzzy functional dependency. 1999 IEEE
- [11] S.Ben Yahia, H.Ounalli, A.Jaoua, An extension of clacical functional dependency: dynamic fuzzy functional dependency. 1999 Elsevier Science Inc.
- [12] Xue Wang, A Simplified Fuzzy Functional Dependency and its Mining Method, International Conference on Fuzzy Information Processing Theories and Applications, March 1-4, 2003, Beijing.
- [13] Zadeh, L. A. Fuzzy sets, Information and control, 1965, 8, 338-35

E-Business Data Warehouse Design and Implementation

Xudong Chen

Computer Department, Jiaying University, MeiZhou 514015, China
chenxd@jyu.edu.cn

ABSTRACT

E-Business have a variety of on-line transaction processing (OLTP) systems and operational database. Data Warehouse is different from operational database. A data warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision-making process. The features of Data Warehouse cause the its design process and strategies to be different from the ones for OLTP Systems. This paper presents a brief description of approaches that address the data warehouse Design and Implementation for E-business.

Keywords: E-Business, Data Warehouse, Data Warehouse Design, OLAP , Dimensional models

1. INTRODUCTION

In the last decade, we witnessed numerous extraordinary business and technology innovations, such as business process reengineering, enterprise resource planning systems, and electronic business^[6]. As transaction databases become more powerful, communication networks grow, and the flow of commerce expands, E-Business have accumulated a variety of on-line transaction processing (OLTP) systems and operational database.

OLTP systems (such as financial, order, work scheduling) which create operational data operational data focuses on transactional functions. This data is part of the corporate infrastructure. It is detailed, nonredundant, and updateable.

However, the rapidly changing market dynamics, competitive pressures, globalization of the commercial markets, reduced profit margins, and other similar factors forced business to review their structures, approaches and strategies^[1].

A data warehouse collects, organizes, and makes data available for the purpose of analysis in order to give management the ability to access and analyze information about its business. This type of data can be called informational data. The systems used to work with informational data are referred to as online analytical processing (OLAP).

OLAP is the technology that enables client applications to efficiently access the data organized by data market and data warehouse. Data warehouse, provide a database organized for OLAP rather than OLTP, can solve OLTP problems.

Data warehousing is one of the most important strategic initiatives in the information systems field. These repositories of data play critical roles in understanding customer behavior in customer to business e-commerce, connecting trading partners along the supply chain, implementing customer relationship management

strategies, and supporting comprehensive performance measurement systems^[7].

2. DATA WAREHOUSING ARCHITECTURE

Bill Inmon defined the term data warehouse: "A (data) warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision-making process."^[4]. Subject-oriented: Data that gives information about a particular subject. Integrated: Data that is gathered into the data warehouse from a variety of sources and merged into a coherent whole. Time-variant: All data in the data warehouse is identified with a particular time period.

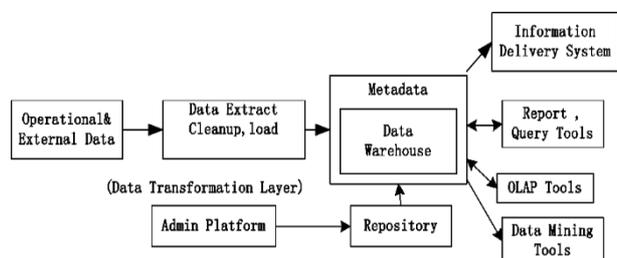


Fig. 1 Data Warehouse Architecture (Refers to Berson 1997)

Data warehouse architecture is based on a relational database management systems server that functions as the central repository for informational data (see Fig.1). Data transformation layer involves a wide range of transformations that have to be applied to source data, for example data quality control and data cleaning, data integration, and conversions that are necessary for adapting data to the DW structures. Metadata: It is data about data that describes the data warehouse, used for building, maintaining, managing, and using the data warehouse. Information delivery system: It distributes data warehouse to other data warehouse and end-user. Data Mining: It is the process of discovering meaningful new correlations, patterns, and trends by mining large amounts of data stored in data warehouse.

3. DATA WAREHOUSE DESIGN AND IMPLEMENT

3.1 Data warehouse design

Data warehouse design is different from traditional OLTP. Kimball,Ralph (1996) proposed Nine-Step Method in the Design of Data Warehouse^[3]: 1. Choosing the subject matter. 2. Deciding what a fact table represents. 3. Identifying and conforming the dimensions. 4. Choosing the facts. 5. Storing precalculations in the fact table. 6. Rounding out the dimension tables. 7. Choosing the duration of the database. 8. The need to track slowly changing dimensions. 9. Deciding the query priorities and the query modes.

Most successful data warehouses design are based on a dimensional model. Dimensional models ^[2,5] represent data with a "cube" structure, making more compatible logical data representation with OLAP data management. The key to build data warehousing is data design. The business users know what data they need and how they want to use it. Focus on the users, determine what data is needed, locate sources for the data, and organize the data in a dimensional model that represents the business needs.

The basic concepts of dimensional modelling include: A fact, dimensions and measures. A fact is a collection of related data items, consisting of measures and context data. It typically represents business items or business transactions. A dimension is a collection of data that describe one business dimension. Dimensions determine the contextual background for the facts; they are the parameters over which we want to perform OLAP. A measure is a numeric attribute of a fact, representing the performance or behaviour of the business relative to the dimensions.

The principal characteristic of a dimensional model is a set of detailed business facts surrounded by multiple dimensions that describe those facts. When realized in a database, the schema for a dimensional model contains a central fact table and multiple dimension tables. A dimensional model may produce a star schema or a snowflake schema.

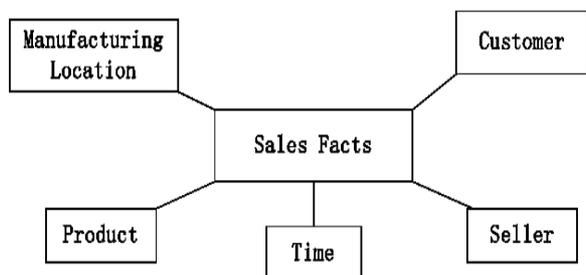


Fig. 2 The star schema example

The star schema is the basic structure for a dimensional model. It has one large fact table and a set of smaller dimensions tables arranged in a radial pattern around the central table.

In this example (see Fig. 2): Sales table is fact table. Time table, customer table, Seller table, Product table and Manufacturing table are dimensional table.

The snowflake schema is the result of decomposing one or more of the dimensions. The many-to-one relationships among sets of attributes of a dimension can separate new dimension tables, forming a hierarchy.

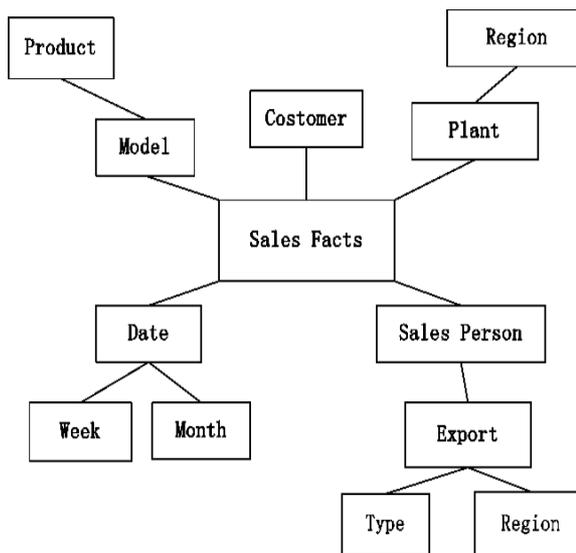


Fig. 3 The snowflake schema example

In this example(see Fig. 3), Sales table is fact table, but one or more dimension tables do not join directly to the fact table but must join through other dimension tables.

Both star and snowflake schemas are dimensional models; the difference is in their physical implementations. Snowflake schemas support ease of dimension maintenance because they are more normalized. Star schemas are easier for direct user access and often support simpler and more efficient queries.

3.2 Data Warehouse Implementation

Data Warehouse implementation must include mechanisms to migrate data into the data warehouse database. This process of data migration is generally referred to as the extraction, transformation, and loading (ETL) process (see Fig. 4).

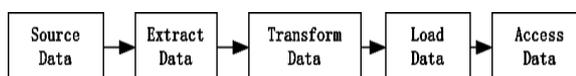


Fig. 4 The Data Warehouse Implementation Process

During the ETL process, data is extracted from an OLTP database, transformed to match the data warehouse schema, and loaded into the data warehouse database. Many data warehouses also incorporate data from non-OLTP systems, such as text files, legacy systems, and spreadsheets; such data also requires extraction, transformation, and loading.

3.2.1 Extract data from source systems

This phase is responsible for extracting data from the source system. During extraction, data may be removed from the source system or a copy made and the original data retained in the source system. It is common to move historical data that accumulates in an operational OLTP system to a data warehouse to maintain OLTP performance and efficiency.

3.2.2 Transform the data

This phase is responsible for data validation, data accuracy, data type conversion, and business rule application. Data Validation: Check that all rows in the fact table match rows in dimension tables to enforce data integrity. Data Accuracy: Ensure that fields contain appropriate values, such as only "off" or "on" in a status field. Data Type Conversion: Ensure that all values for a specified field are stored the same way in the data warehouse regardless of how they were stored in the source system. Business Rule Application: Ensure that the rules of the business are enforced on the data stored in the warehouse.

3.2.3 Load data to Data warehouse

This phase is responsible for loading transformed data into the data warehouse database. Data warehouses are usually updated periodically rather than continuously, and large numbers of records are often loaded to multiple tables in a single data load. The data warehouse is often taken offline during update operations.

3.2.4 User and application access the data

This phase rely on a suite of tools. The best way to choose this suite includes the definition of different types of access to the data and selecting the best tool for that kind of access. In general, access types include: Reporting, Ranking, Multivariable analysis, Time series analysis, Data visualization, graphing, charting, and pivoting, Statistical analysis, Artificial intelligence technique, OLAP tool, data mining tool, etc. Many tools and applications can be used to access warehouse data.

4. CONCLUSION

As transaction databases become more powerful, communication networks grow, and the flow of commerce expands, E-Business have accumulated a variety of on-line transaction processing (OLTP) systems and operational database. Data Warehouse is different from operational database. Data warehouses have been used to support business decision makers. A data warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision-making process.

The key to build data warehousing is data design. The business users know what data they need and how they want to use it. Focus on the users, determine what data is needed, locate sources for the data, and organize the data in a dimensional model that represents the business needs. Dimensional modeling is the foundation of data warehouse design. A dimensional model may produce a star schema or a snowflake schema. the schema for a dimensional model contains a central fact table and multiple dimension tables.

Data Warehouse implementation include: extracting, transforming, and loading the data into the data warehouse, creating the OLAP and data mining analytical applications, developing end-user tools.

REFERENCES

- [1] Berson, A., S.J. Smith, Data Warehousing, Data Mining, and OLAP, McGraw-Hill, 1997.
- [2] Chaudhuri, S., U. Dayal., "Database Technology for Decision Support Systems", Computer, Vol 34, No12, pp48-65, 2001.
- [3] Kimball, R., The Data Warehouse Toolkit, Wiley, New York, 1996.
- [4] Inmon, W.H., Building the Data Warehouse, Wiley, New York, 1992.
- [5] Pedersen, T.B., C.S. Jensen, "Multidimensional Database Technology", Computer, Vol 34, No12, pp40-46, 2001.
- [6] Markus, M.L., "Paradigm Shifts-E-business and Business/Systems Integration", Communications of the Association for Information Systems, Vol 4, pp1-44, 2000.
- [7] Watson, H.J., "Recent Developments in Data Warehousing", Communications of the Association for Information Systems, Vol 8, pp1-25, 2001.

Forecasting the Consumption for Electricity in Taiwan

Hsiaotien Pao¹, Tenpao Lee²

¹ Department of Management Science, National Chiao Tung University, Taiwan, China
htpao@cc.nctu.edu.tw

² Department of Commerce, Niagara University, New York, USA
tlee@niagara.edu

ABSTRACT

This paper use linear regression and non-linear artificial neural network (ANN) model to analyze how the four economic factors: national income (NI), population (POP), gross of domestic production (GDP), and consumer price index (CPI), affect Taiwan's electricity consumption, furthermore, develop an economic forecasting model. Both models agree with that POP and NI are of the most influence on electricity consumption, whereas GDP of the least. Then, we compare the out-of-sample forecasting capabilities of the two models. The comparing result indicates that the linear model is obviously of higher bias value than that of ANN model, and of weaker ability of forecasting capability on peaks or bottoms. This probably results from: 1) linear regression model is built on the logarithm function of electricity consumption, and ANN is built on the original data; 2) ANN model is capable of catching sophisticated non-linear integrating effects. Consequently, ANN model is the more appropriate between the two to be applied to building an economic forecasting model of Taiwan's electricity consumption.

Keywords: electricity consumption, artificial neural network, linear regression

1. INTRODUCTION

Modeling electrical demand and energy consumption is usually based on historical consumption and the relationship of this consumption to other relevant variables, such as: economic, demographic, climatic, energy price, etc. Multivariate modeling along with cointegration techniques or regression analysis were used in a number of studies on different countries [1-7] to investigate the influence of different determinants on energy consumption. Recently, some studies address to compare forecasting performance for energy consumption using different models on different countries [8-10]. Variables affecting demand and energy consumption may vary from one region to another. A model developed for one region may not be appropriate for another region. Electrical consumption models are required for a variety of utility activities. Therefore, model should be developed for different region for efficient planning and organization.

Taiwan's energy consumption rises sharply from 52.01 million kiloliters of oil in 1990 to 103.42 million kiloliters of oil in 2003 because of rapid economy growth and higher living standard. Among the energy forms consumed, petroleum took up 44.2% in 1990 to 39% in 2003; coal 12.8% to 10.7%; natural gas & liquid natural gas 3% to 2.3%, and electricity 40% to 48%. Electricity consumption takes up almost 50% of the total final energy consumption in 2003. The main object of this study is to analyze and discuss Taiwan's electricity consumption. We focus on the effect brought by four economic factors: NI, POP, GDP, and CPI, on electricity consumption (ELEC) for Taiwan. Moreover, to build a linear and a non-linear economy forecasting models which could be useful to relating government authorities for controlling electrical energy supply.

The rest of this paper is organized as follows: section 2 records the results of data analysis and review. Section 3 presents the empirical results and comparing results of linear and nonlinear models, whereas the last section briefs our findings and concludes.

2. DATA ANALYSIS

The data applied here are 156 monthly data recorded January 1990 through December 2002. The first 132 data recorded 1990 through 2000 are studied to build the forecasting model, while the last 24 data recorded 2001 through 2002 are applied to testify the model. ELEC and the four economic factors: NI, POP, GDP, and CPI, are collected from AREMOS database supervised by the Education Ministry Taiwan. The factor NI has rather high linear relativity with POP (0.99) and CPI (0.97), while POP reacts the same to CPI (0.96). The scatter plot of ELEC to four economic factors exhibit logarithmic trend. So, Log (ELEC) is considered dependent variable in this research. Fig. 1 shows the electricity consumed annually from 1990 to 2002. The consumption gradually increases to hit the peak in summertime, and decreases to the bottom in wintertime.

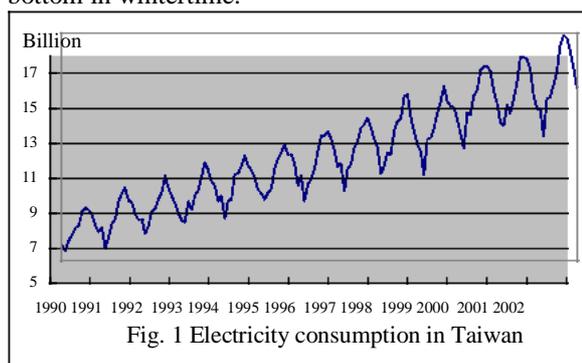


Fig. 1 Electricity consumption in Taiwan

3. EMPIRICAL RESULTS

The following five linear and one nonlinear models are used to find how POP, NI, GDP, and CPI affect Taiwan’s electricity consumption, furthermore to build a forecasting model for the issue.

3.1 Multiple linear regression models

This section analyzes and discusses the effect that the four economic factors have on electricity consumption with linear regression model, and develops an economic forecasting model. Because of the linear relationships between logarithm function of electricity consumption and economic factors, this paper use the Log (ELEC) as dependent variable. Taking multicollinearity into account, the experimental formulas are extracted after deleting NI and CPI variables selected by the models with statistics of Mallows C(p) [11], Adj-R², and VIF.

- Model 1: $LELEC_t = a_1 + b_1TEMP_t + c_1GDP_t + e_{1t}$
 - Model 2: $LELEC_t = a_2 + b_2TEMP_t + c_2CPI_t + e_{2t}$
 - Model 3: $LELEC_t = a_3 + b_3TEMP_t + c_3POP_t + e_{3t}$
 - Model 4: $LELEC_t = a_4 + b_4TEMP_t + c_4NI_t + e_{4t}$
 - Model 5: $LELEC_t = a_5 + b_5TEMP_t + c_5NI_t + d_5GDP_t + e_{5t}$
- where $LELEC = \text{Log (ELEC)}$.

Model 1- 4 present how the four economic factors affect electricity consumption under controlled the temperature,

respectively. Table 1 indicates that POP and NI affect Taiwan’s electricity consumption the most reaching 95.18% and 95.17% respectively followed by CPI (92.48%), while GDP the least (59.25%). The normality test for residuals of each model is recorded in Table 1. Though GDP and CPI are both significant indicators to a country’s economy, they don’t affect Taiwan’s electricity consumption as much as the others. This could result from Taiwan government’s enforcement on electricity saving and economizing policy.

Model 5, the best economic forecasting model for electricity consumption, is formulated after deleting NI and CPI variables selected by the statistics of Mallows C(p), Adj-R², and VIF. The procedure is practiced to prevent multicollinearity. Compared with model 4, model 5 shows that GDP has significant effect on Taiwan’s electricity consumption, but it only surpasses model 4 by 0.34 (=0.9551 – 0.9517). The last row in Table 1 present the estimated coefficient values of standardized regression model. It shows the most crucial factors affecting Taiwan’s electricity consumption, according to priority, are NI (0.803), TEMP (0.381), and GDP (0.082). The result agrees with those of model 1 to 4. The Adj-R² of model 5 reaches 0.9551, and passes the normality test for residuals ($p\text{-val}=0.1438$). Below is the formula of model 5:

$$LELEC=14.639+0.024TEMP+1.47*10^6 NI+5.44*10^8 GDP$$

Table 1 Coefficients using regression and standardized regression

		Intercept	TEMP	GDP	CPI	M	NI	Adj-R ²	Normality (p-val)	Size
Model 1	Estimate	14.462* (120.71)	0.028* (7.77)	4.07E-7* (10.92)				0.5925	0.0060	132
	Standardized	0	0.434	0.610						
Model 2	Estimate	13.134* (176.36)	0.022* (14.42)		0.027* (34.88)			0.9248	0.3100	132
	Standardized	0	0.349		0.845					
Model 3	Estimate	8.019* (47.27)	0.023* (18.29)			3.59E-4* (44.38)		0.9518	0.0379	132
	Standardized	0	0.355			0.860				
Model 4	Estimate	14.726* (466.41)	0.024* (19.69)				1.57E-6* (44.33)	0.9517	0.0197	132
	Standardized	0	0.381				0.857			
Model 5	Estimate	14.639* (364.79)	0.024* (20.45)	5.44E-8* (3.30)			1.47E-6* (32.30)	0.9551	0.1438	132
	Standardized	0	0.381	0.082			0.803			

* Indicates statistical significance at the 0.05 level. t statistics are in the parentheses.

The model residuals are amplified by exponential transformation because we are fitting the regression model to transformed data (Log (ELEC)) and then must convert prediction for logarithmic back to untransformed units. So, we build a non-linear ANN model on the original value of electricity consumption.

3.2 ANN models

The artificial neural network (ANN) consists of an input layer, an output layer and one or more intervening layers also referred to as hidden layers. Each layer consists of multiple neurons that are connected to neurons in adjacent layers. The connection weights and node biases are the model parameters.

To use an ANN model for forecasting, forecasters must first build it. The model building process is called the network training or learning. Usually, in applications of ANNs, total available data are split into a training set and a test set. The training set is used to build the network model and then the forecasting ability of the network is evaluated from the test set. In this paper, the 132 monthly data recorded 1990 through 2000 are training samples, while the 24 monthly data recorded 2001 through 2002 are testing samples. POP, NI, GDP, CPI, and TEMP are the network input values, and electricity consumption is network output value. More than 60 experiments are performed to determine the best combination of the learning rate, momentum, number of hidden layers, number of hidden nodes, learning rule, and transfer

$$S_i = \frac{\partial \hat{Y}}{\partial X_i} = \sum_{t=1}^h \frac{\partial \hat{Y}}{\partial NET} \frac{\partial NET}{\partial Z_t} \frac{\partial Z_t}{\partial NET} \frac{\partial NET}{\partial X_i} = \sum_{t=1}^h [f'(NET) v_t f'(NET) w_{it}] \quad (1)$$

The absolute sensitivity value of each input variable represents its relative effect with electricity consumption. Crucial factors affecting electricity consumption are, by priority, NI, POP, TEMP, CPI and GDP, and their absolute sensitivities are 15.142, 7.696, 3.068, 2.344 and 0.989 (see Table 2). The result agrees with that of regression model on which the economy indicators, GDP and CPI, are not the key factors to Taiwan's electricity consumption, but POP and NI.

Table 2 The sensitivities of electricity consumption to five economic factors

	TEMP	GDP	CPI	POP	NI
Sensitivity	3.068	0.989	-2.344	7.696	15.142

Once the two models to predict monthly electricity consumption are developed, we empirically examine the relative effectiveness of the models in predicting electricity consumption using the data from January 2001 to December 2002.

3.3 Out-of-sample forecasting performance results

We use three statistics: RMSE, MAE, MAPE, and scatter diagram to see the out-of-sample ability of the linear and the non-linear models. Table 3 shows the three statistic values fall sharply from 1509559.37, 1342149.39, and 8.60% of linear model to 635378.69, 460738.5, and 3.29% of ANN model. It is because of: 1) the dependent variable of regression model is the logarithm value of electric consumption, while the artificial neural network model trains the original value of electricity consumption; 2) the ANN is capable of catching sophisticated non-linear integrating effect. Fig. 6 presents the forecasting values of two forecasting models and ELEC actual value. It clearly shows that the ANN forecast value is closer to the ELEC actual value, and regression model is weaker on foretelling peaks and bottoms. This result proves that the ANN is more appropriate to be applied to build an economic forecasting model for Taiwan's electricity consumption.

function to utilize.

Throughout the training, the NeuralWare utility, 'SAVEBEST' is used to monitor and save the lowest root mean square (RMS) error value from the testing set. The best RMS error result is obtained using a learning rate of 0.2, a momentum of 0.1, and 6 neurons in a single hidden layer that use the generalized delta learning rule and a sigmoid transfer function. The architecture of the best network contains 5 input layer neurons, 6 hidden neurons, and 1 output layer neuron (5:6:1 architecture). The estimations of connection weights w_{ij} and v_i are obtained, then apply them to formula (1) to calculate the sensitivities \hat{S}_i (Hwang, Choi, Oh, & Marks [12]).

Table 3 Comparing forecasting measurement errors

	RMSE	MAE	MAPE
Regression	1509559.37	1342149.39	8.60%
ANN	635378.69	460738.50	3.19%

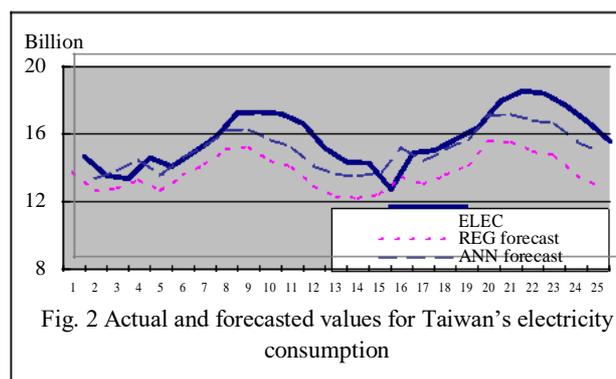


Fig. 2 Actual and forecasted values for Taiwan's electricity consumption

4. CONCLUSION

By adopting linear regression and non-linear ANN models, surprisingly, we find that economy indicators, GDP and CPI, have less effect on Taiwan's electricity consumption than POP and NI. By ANN model, we obtain the crucial factors are, by priority, NI, POP, TEMP, CPI and GDP. Three factors: NI, TEMP, GDP are contained in the best linear economic forecasting model, after excluding the multicollinearity among five factors. By comparing three statistics values: RMSE, MAE, MAPE, and scatter diagram for out-of-sample of two models, ANN's forecasting capability is higher than that of the regression model, and regression model is weaker on foretelling peaks and bottoms. To sum up, the artificial neural network model is more appropriate between the two to help us build the economy-forecasting model of Taiwan's electricity consumption. Furthermore, it is possible to use linear and nonlinear hybrid model and univariate time series to forecast electricity consumption in Taiwan.

ACKNOWLEDGEMENT

This research was supported in part by the National Science Council under Grant NSC 92-2422-H-009-0300 in Taiwan, China.

REFERENCES

- [1] Glasure, Y.U., Lee, A.R. Cointegration, error-correction, and the relationship between GDP and electricity: the case of South Korea and Singapore. *Resource and Electricity Economies* 1997;20:17-25.
- [2] Shiu A, Lam PL. Electricity consumption and economic growth in China. *Energy Policy* 2004;32:47-54.
- [3] Soytaş U, Sari R. Energy consumption and GDP: causality relationship in G-7 countries and emerging markets. *Energy Economics* 2003;25:33-37.
- [4] Yang HY. A note on the causal relationship between electricity and GDP in Taiwan. *Energy Economics*. 2000;22:309-317.
- [5] Yan YY. Climate and residential electricity consumption in Hong Kong. *Energy* 1998;23(1):17-20.
- [6] Ranjan M, Jain VK. Modeling of electrical energy consumption in Delhi. *Energy* 1999;24:351-361.
- [7] Egelioglu F, Mohamad AA, Guven H. Economic variables and electricity consumption in Northern Cyprus. *Energy* 2001;26:355-362.
- [8] Darbellay GA, Slama M. Forecasting the short-term demand for electricity: do neural networks stand a better chance. *International Journal of Forecasting* 2000;16:71-83.
- [9] Saab S, Badr E, Nasr G. Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon. *Energy* 2001;26:1-14.
- [10] Fatai K, Oxley L, Scrimgeour FG. Modeling and forecasting the demand for electricity in New Zealand: a comparison of alternative approaches. *The energy Journal* 2003;24(1):75-102.
- [11] Mallows CP. Some Comments on C(p). *Technometrics* 1973;15:661-675.
- [12] Hwang, J.N., Choi, J.J., Oh, S., & Marks, R.J. Query Based Learning Applied to Partially Trained Multilayer Perceptron. *IEEE T-NN* 1991;2(1):131-136.

Logistics Forecasting Using Improved Fuzzy Neural Networks System

Jianyu Zhang¹, Xuelei Pu¹, Sheng Li¹, Dan Yang²

¹School of Business Administration, Southwestern University of Finance & Economics,
Chengdu 610074, China

²MBA Education Center, Southwestern University of Finance & Economics,
Chengdu 610074, China

ABSTRACT

In this paper, we proposed and trained a fuzzy neural network system to estimate future logistics demand. The structure of neural network in the system is similar to that of BP network, except that here the nonlinear sigmoid functions in the networks are replaced by fuzzy reasoning process and wavelet functions respectively. Moreover, the trained network system is put into practical logistics demand forecasting. The experimental results show that it has good properties such as a fast convergence, high precision and strong function approximation ability and is good at predicting future logistics amount.

Keywords: logistics demand, forecasting, fuzzy logic, wavelet neural networks

1. INTRODUCTION

In reality, it is significant to estimate future logistics amount as to a logistics manager in a company. Accurate prediction of logistics amount will help the managers to make decision more right. The indefinite nature of the market means that there is a significant need for accurate forecasts of future business, both short- and long-term. Hence, managers, researchers and policy makers have recognized the necessity of accurate forecast of logistics demand. They have tried to apply various kinds of mathematical techniques to reflect the quantitative relationship between a specific demand and its impact factors. On the other hand, another approach to set up a relationship model for the dependent variable only based on its past data or performance. The relationship model is then used to estimate future data.

Previous relationship modeling methods have been largely based on time-series models and multivariate regression analyses. In summary, there are many forecast approaches in the world, each of which has remarkable superiority and unavoidable weakness.

Back-propagation (BP) neural network can establish function approximation for specific input and output relationship without a given model. Nevertheless, the traditional BP network has some weaknesses that are easy to relapse into local minimum point and its computation convergence speed is somewhat slower, which affect the enhancement of reliability and accuracy of prediction model. Thus, we put forward a novel prediction model for the logistics demand based on the combination of fuzzy logic and neural network. We combine fuzzy reasoning theory with neural network in order to improve the precision and convergence speed of the prediction model. So, a neural network-driven fuzzy reasoning system is proposed on the basis of improved Takagi-Sugeno reasoning model. The system consists of $P+1$ neural networks, the

networks $NN_1 \sim NN_P$ denote the functions in the conclusion parts of the P rules respectively, and this type of networks are called expert networks; NN_{P+1} is used to calculate the fitness of each rule corresponding to input vector, this type of networks is called gate network. Thus, the topology structure of the neural network-driven fuzzy reasoning system is determined. We think that it is feasible to model the relationship among the time-series historic data of logistics demand using fuzzy neural networks system. Namely, making no assumption about other impact factors, time-series forecast models based on fuzzy neural networks can use historical data as the input and output variables to form a univariate or multivariate mathematical function to reflect the relationship among these statistical data.

2. FUZZY NEURAL NETWORKS SYSTEM

2.1 Introduction to Fuzzy Neural Networks

In recent years, artificial neural network, a commonly used nonlinear function approximation tool, has shown huge advantages in forecasting, pattern identification, optimization technique and signal processing for its good properties such as nonlinear, flexible and valid self-organization study et al. Back-propagation (BP) neural network, a typical case of neural networks, is used most widely and is more mature than other networks. Unlike the classic mathematic methods, BP networks can approximate the specific input and output relationship without a certain model. Therefore, we usually attempt to set up a BP ANN model for logistics demand prediction. Nevertheless, traditional BP network has some weaknesses that it is easy to relapse into local minimum point and computation convergence speed is somewhat slower, which affect the enhancement of reliability and precision of prediction model. So, we intend to combine fuzzy reasoning theory with neural networks in order to improve the precision and convergence speed of simulation model.

Thus, a neural network-driven fuzzy reasoning system is proposed on the basis of improved Takagi-Sugeno reasoning model. The system consists of $P+1$ neural networks, the networks $NN_1 \sim NN_P$ denote the functions in the conclusion parts of the P rules respectively, and NN_{P+1} is used to calculate the fitness of each rule corresponding to input vector.

As there are a good deal of practical application of fuzzy neural networks are reported up to now, its superiority for solving complicated non-linear problems has been proved by these researches and taken seriously by scholars gradually.

In this paper, we intend to approximate and fit the functional relationship between the among the statistic data of logistics demand using the neural network driven fuzzy reasoning system abovementioned, and corresponding learning algorithm of fuzzy neural networks system is proposed there. Thus, a prediction model of logistics demand based on fuzzy neural networks system is hereby founded and is put into practice.

2.2 Fuzzy Neural Networks System

As to a fuzzy system with n -input and single output, a regular style of fuzzy rules is:

If x_1 is A_1 , x_2 is A_2 , ..., x_n is A_n , then y is B , where A_1, A_2, \dots, A_n and B is fuzzy subsets. In the conclusion part, the model substituting fuzzy set B with a function is commonly called Takagi-Sugeno (TS) model.

If x_1 is A_1 , x_2 is A_2 , ..., x_n is A_n , then $y=f_j(x)$, where $f_j(x)$ is a linear combination of input variables, namely

$$f_j(x) = c_1^j x_1 + c_2^j x_2 + \dots + c_n^j x_n + c_j^{n+1} \quad (1)$$

The model divide input space into linear spaces because the input variables are independent each other. When we use the two models abovementioned to divide input space into nonlinear spaces, it is necessary to have an elaborate division at first. But, it will result in the rapid increase of the number of fuzzy rules. To avoid the happening of such a situation, it is convenient to adopt the number of following model:

$$\text{If } X \in P_j, \text{ then } y = f_j(X) \quad (2)$$

Where $X = (x_1, x_2, \dots, x_n)$, P_j is a partial space divided from input space. In the formula (2), the membership function cannot be determined as formula independently, so we only acquire the joint membership function in the condition part using neural networks, similarly, the function in the conclusion part can also be expressed by neural networks. Thus, a fuzzy system based on neural networks is proposed and its structure is shown in Fig 1.

The system consists of $P+1$ neural networks, where $NN_1 \sim NN_P$ denote the functions in the conclusions of the P rules, NN_{P+1} is used to calculate the fitness of each rule corresponding to input vector. The output of the

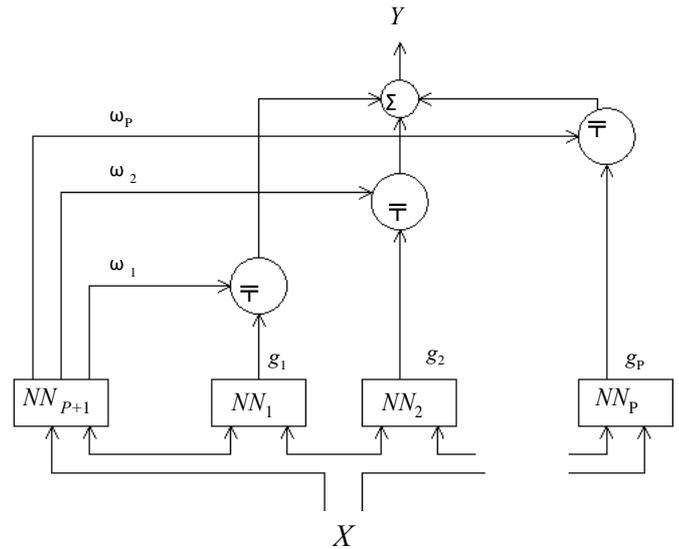


Fig.1 A fuzzy neural networks system

fuzzy system can be calculated by the following formula

$$Y = \sum_{j=1}^P \omega_j g_j \quad (3)$$

Where g_j is the output value of the network NN_j and ω_j is its corresponding fitness. The system is called an NN-driven fuzzy reasoning system.

We stipulate three kind of function of the fuzzy system as follows:

- 1) Determine the number of fuzzy rules by the use of K-means clustering method.
- 2) Design and train the neural network NN_{P+1} to calculate the fitness of each rule corresponding to input vector.
- 3) Design and train the neural network NN_j to denote the nonlinear function in the conclusion part in the fuzzy system.

The steps of establishing and training the fuzzy system are as follows:

- Step 1. Collect training sample for the fuzzy system.
- Step 2. Use K-means method to cluster input vectors, and each cluster corresponds to a rule. Because the samples are clustered into P teams, there are P fuzzy rules for the fuzzy reasoning system.

Step 3. Train the network NN_{P+1} , which has n inputs and P outputs in its structure. The training samples can be constructed using the following method:

If the training sample X is clustered into the S -th team, we have:

$$\omega_j^i = \begin{cases} on & j = S \\ off & j \neq S \end{cases} \quad j = 1, 2, \dots, P \quad (4)$$

Namely,

$$W^i = (\omega_1^i, \omega_2^i, \dots, \omega_S^i, \omega_{S+1}^i, \dots, \omega_P^i)^T$$

Where on/off correspond to 1 or 0 respectively. But the

output of the activation function in the neural networks is not absolutely 1 or 0. Here we replace them with 0.9/0.1, which can accelerate the training process of networks. Thus, we can construct training sample (X_i, W^i) for the network NN_{P+1} . After a certain number of training, the final fitness vector $W = (\omega_1, \omega_2, \dots, \omega_p)^T$ for each network NN_j ($j=1, \dots, P$) is gained.

Step 4. Train the networks $NN_1 \sim NN_P$. Assume that NN_s ($s=1, \dots, P$) corresponds to the S -th rule, thus, all of the samples in the S -th cluster is the training data for the network NN_s . Then, back-propagation algorithm is applied to train the networks $NN_1 \sim NN_P$.

2.3 Applying Wavelet Transformation and Wavelet Network in the Networks NN_j

In nature, wavelet transform is a kind of integral transform among various parameters:

$$\omega(a, b) = \int_{-\infty}^{+\infty} f(t) \cdot h(a^{-1}(t-b)) dt \tag{5}$$

where $f(t)$ is a function with compact support set, $h(a, b, t)$ is called wavelet such that

$$h(a, b, t) = \frac{1}{\sqrt{|a|}} h_{basic}\left(\frac{t-b}{a}\right)$$

Here $h_{basic}(t)$ is called basic wavelet or mother wavelet, $\frac{1}{\sqrt{|a|}}$ is the coefficient of normalization, and

a, b are dilation and translation factor of $h(a, b, t)$. As to a signal or function $f(t)$, the identification of its local structure can be conducted by adjusting those two parameters to change the size and place of wavelet bases window to realize.

From the wavelet theory and latest research reports, we know that an arbitrary function $f(x) \in L^2(R^n)$ can be approximated closely as

$$f(x) \approx \sum_{j=1}^n \omega_j \varphi\left(\frac{x-b_j}{a_j}\right) \tag{6}$$

Formula (6) is a general mathematic description for wavelet network, where n is the number of network nodes, $\varphi(\bullet)$ is basic wavelet, b_j is translation factor, a_j is dilation factor and ω_j is link weight of the network.

Wavelet networks is a type of novel feedforward network on the basis of wavelet analysis theory. It utilizes wavelet space as the characteristic space of style identification, and realizes the feature identification by the weighed sum of the inner product for basic wavelet and signal vector. Because it is a combination of time-frequency localization characteristic of wavelet transform and self-study function of artificial neural network, it owns such capabilities as valid function approximating and error bearing. The function relationship of formula (6) can be

realized by using linear superposition of nonlinear basic wavelet, namely, we can approximate formula (6) using the limited terms of wavelet series.

Assume that there is a three-layer wavelet network whose input vector is $X=(x_1, x_2, \dots, x_n)$ and output vector is $Y=(y_1, y_2, \dots, y_m)$ and the number of its neurons is n, q, m , respectively and the link weights of the network are ω_{ij} and ω_{jk} . Thus, the nonlinear function $f(\bullet)$ in formula (6) can be fitted using basic wavelet $h(a, b, t)$. So, we have

$$f(\bullet) = \sum_{j=1}^q \omega_{jk} h\left(\frac{\sum_{i=1}^n (\omega_{ij} x_i) - b_j}{a_j}\right) \tag{7}$$

where a_j, b_j are the dilation factor and translation factor respectively, and q is the number of wavelons in the hidden layer.

Put the input and output vector of P samples into the network, and calculate the output values and the system error of the network. To assess the approximation results, we select the least squares objective function as a figure of merit, namely

$$E = \frac{1}{2} \sum_{l=1}^P \sum_{k=1}^m (d_{lk} - y_{lk})^2 \tag{8}$$

where y_{lk} is the actual value of l th sample on the k th output node in the network, and d_{lk} is the ideal value of the k th output node accordingly.

In the training process of wavelet network, we select Morlet mother wavelet as the basic wavelet approximating the nonlinear function such that

$$h(t) = \cos(1.75t) \exp\left(-\frac{t^2}{2}\right) \tag{9}$$

The steps of wavelet network training are as follows:

Step 1. Initializing the parameters of the wavelet network. Using random generator, the link weights of the network ω_{ij}, ω_{jk} and the dilation factor a_j and translation factor b_j of basic wavelets are evaluated with random number.

Step 2. Project study samples data into the network.

Step 3. Carry out the self-study of network and gain the output vector and error of network.

Step 4. Calculate the instantaneous gradient vector.

Step 5. Carry out error back-propagation and revise the parameters of the networks.

Step 6. While the absolute value of the error is less than the precision, stop the training process. Otherwise, turn to Step 3.

Step 7. Put the trained wavelet network into the practice.

The functions in the conclusion part of the fuzzy reasoning system are all expressed by wavelet neural

networks NN_j ($j=1, \dots, P$).

3. FUZZY NEURAL NETWORKS SYSTEM FOR LOGISTICS DEMAND FORECATING

In the work here, this novel networks system is applied to model logistics forecasting. What we should do firstly is to decide the variables denoting the input-output relationship for it. Generally, whichever networks can be used in time-series prediction if it can accept actual input and brings out actual output. If the prediction objective is a future point in a single time-series, it is feasible to use the former $n-1$ points before this point as the historic data to estimate its value. In a time series, every n neighboring point can be used

as training samples. Here, the proposed fuzzy neural networks system is used as a time-series prediction model for logistics demand. Namely, every $n-1$ neighboring point can be used as training sample and put into the system to estimate the value of the n -th point.

So, the training sample data and testing data of fuzzy neural networks for logistics forecasting are collected and constructed. In the process of practical application, for the difference in the dimension and magnitude of sample data, all inputs are scaled to lied in $[0, 1]$ range in advance using reasonable estimation of the maximum and minimum values for each input.

Table 1. Comparison between prediction and the actual value of rail freight amount in China

Year	1989	1990	1991	1992	1993	1994	1995	1996	1997
Actual	151489	150681	152893	157627	162663	163093	165855	168803	169734
Prediction	151390	150540	157512	157432	160565	162187	166238	169126	170163

We apply the trained networks system in the study of actual logistics forecasting and by comparison with historical statistics data to test its properties. The experimental results are contrast to that of actual data and shown in Table 1. The experimental results show that although there a certain error in the prediction values to some extents, the simulation model can forecast logistics demand properly and accurately.

4. CONCLUSIONS

In this paper, we proposed and trained an improved fuzzy neural networks system with the data of past logistics demand data. The structure of the networks in the system is similar to that of BP network, except that here the nonlinear sigmoid functions in the networks are replaced by fuzzy reasoning process and wavelet functions respectively. Moreover, the trained network is put into practical logistics demand forecasting. The experimental results show that the fuzzy neural network has good properties such as a fast convergence, high precision and strong function approximation ability. Furthermore, it also demonstrates that the fuzzy neural network system is good at predicting future logistics amount because of its non-strict requirements for input

variables and not needing plenty of sample data, which ensure that it is suitable for the actual forecasting and outperforms other classic prediction methods and is significant to in applied in the planning of logistics system in both a company and a city.

REFERENCES

- [1] Parlos, A. G., "An accelerated learning algorithm for multiplayer perceptron networks", *IEEE Transaction on Neural Networks*, Vol. 5, pp. 493-497, 1994.
- [2] Tong, R. M., "A control engineering review of fuzzy systems", *Automatic*, Vol. 13, No. 6, pp 559-569, 1997.
- [3] Kim, K. H., Park, J. K., et al., "Implementation of hybrid short-term forecasting system using network and fuzzy expert systems", *IEEE Transaction on Power Systems*, Vol. 10, No. 3, pp1534-1539, 1995.
- [4] Cybenko, G., "Approximation by superpositions of a sigmoidal function", *Math. Control Signals Syst.*, Vol. 2, pp 303-314, 1989.
- [5] Wang, L. X., Mendel J. M., "Fuzzy basis functions, universal approximation and orthogonal least-squares learning", *IEEE Transaction on Neural Networks*, No. 3, pp807-814, 1992.

Maximising Technology Efficiencies for SMEs Using Computer Intelligence

Adrian Collins

University of Wollongong, Australia
adrian@uow.edu.au

ABSTRACT

This paper is part of a series which present elements of our research into technology adoption and usability problems faced by small businesses (SME's) in Australia. We discuss our approach to research in the small business area, the importance of aggregating small businesses into vertical technology sectors and the usability improvements which can be expected from this collaborative approach. In this paper we focus on information retrieval requirements and the related usability problems which confront small businesses. We discuss adoption and usability issues, requirements and risks using a business case study, then introduce experiments which it is hoped will deliver efficiency improvements to information retrieval applications for small businesses.

Keywords; Small Business, collaborative approach, content ranking

1. INTRODUCTION

Our goal is to help SME's find innovative and novel ways to utilise technology. We believe SME's can significantly benefit from a selective adoption of technologies to improve efficiencies for staff by improving communication and information sharing facilities with other organisations. Technology adoption can be extremely cost effective for small business and provide a genuine return on short term investment. Our research aim is to test the viability of improving access to technology for SME's by using a collaborative approach to investment, acquisition and adoption of information and communication technology. It is hoped that this approach will allow individual SME's to spread the cost and risk of eBusiness adoption and use by entering into vertical market collectives. Typically, around 25% of Australian businesses are making some use of electronic business (eBusiness) to reduce the cost of doing business or improve profit potential. Recent research into the local region has found that eBusiness adoption by SMEs is as low as 6%. The major reasons for not adopting eBusiness technology by SMEs:

- Not relevant to the business profile
- Technology is too complicated
- The organisation does not have the resources to implement the technology

2. AGGREGATING SME'S INTO A VERTICAL SECTOR

In Australia, many small businesses are focused on supporting a single, larger business entity or market area. In this case, the SME will have implemented the necessary technology and infrastructure to complete the business process in the most profitable way but lacks the resources to further expand. Many organisations would like to grow their business but due to a lack of resources (finances, staff, expertise etc.) cannot take advantage of growth opportunities. This lack of resources also acts as

a barrier to adopting technology which allows them to expand and service new markets. We propose a collaborative approach in which small business stakeholders can pool resources and construct an eBusiness facility that all share. This approach fosters closer business relationships and allows individual SMEs to pool resources by aligning organisations with similar attributes into vertical markets (technology wise). Once organisations with similar attributes are aligned, benefits can be derived by developing and implementing common software solutions which both reduce the cost of doing business and maximise operational efficiencies within the organisation.

An example of organisations with similar attributes is the mercantile agent market, which has more than 1000 small, independent organisations operating in Australia. A mercantile agent acts on a client's behalf and represents persons and firms by collecting overdue debt and by identifying individuals and organisations implicated in credit fraud and theft. When these types of small organisations are analyzed, we observe that while they are physically separate small organisations involved in a specific business task, the area and scope of their business overlaps that of other small businesses involved in similar operational activities. In the main, mercantile agencies perform services such as:

- Collection: make demands for payment of overdue debts
- Investigation: obtain information and report on the personal character, occupation, activities or actions of persons or businesses
- Process serving: deliver court issued documents for a variety of reasons including debt recovery, family law proceedings, civil and criminal litigation.
- Repossession: act for principals when financial contracts (leases, bills of sale, hire purchase agreements etc.) have been broken. Seeking payment of arrears or the balance of a loan facility

or else to recover the goods on behalf of the principal.

While the operational activities are different among the organisations within the vertical market, there is a significant overlap in many areas. Agents share the same types of clients, acting for banks, financiers, lawyers, insurers, the general business community or individuals. In addition, they share many similar information processing requirements. Document Management Systems (DMS), precedents, document archival, mail merging, diary systems, backup and disaster recovery.

We hypothesize that small businesses can save considerable capital investment dollars by aggregating their processing requirements and benefit by collaboration [1]. These benefits are achieved via the use of portal technology which has been shown to result in a number of business and technical benefits:

- Increased productivity and efficiency
- Provides organisations with the ability to present a central point of access to data and organizational documentation
- Facilitates mobile and pervasive device access
- Increases marketing opportunities
- Provides access to component based solutions

3. CASE STUDY

A mercantile agency typically follows the business model shown in figure 1. Currently 10% of matters are handled outside the local court domain. The business sees this as a major growth area, but the trade off is being able to maximise business efficiencies and promote a larger client base against maintaining a small number of personnel. In order to maximise business efficiencies, the organisation invests in technology, deciding that technology adoption will grow the business exponentially, while investment in staff with only grow the business according to the number of additional staff members.

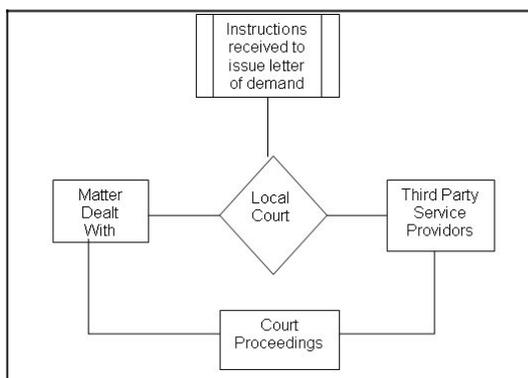


Figure 1.

When the matter is to be dealt with (handled locally), the staff communicate with the debtor and local court on the clients behalf, this process may also include

communication with solicitors on both sides. However, when the debtor resides in another locality, the matter may need to be heard in that court locality. In this case, service providers who may act on the mercantile agencies behalf must be located, fees negotiated, and paid before action can commence in that locality.

This presents a major problem for the organisation, because there is no recognised registry of agents. The organisation must locate potential service providers using either regional telephone directories or the internet. We further discuss this problem in section 3.2 and propose a possible solution.

4. PROCESSING SCENARIOS

4.1 Document Management System

The case study organisation also maintains a DMS (Document Management System) with standard processing requirements, ie. Creation and maintenance of documents which typically relate to judgment matters in a default of payment claim:

- Letter of Demand
- Issue of Statement of Claim
- Service of Statement of Claim
- Default Judgment
- Writ of Execution or Warrant
- Examination Summons
- Matter action record

Many documents may be created for a single matter, and these documents need to be retained in hard copy (and online) for a number of years (7) by law. Given this requirement, an efficient method of retrieving specific documents was identified in the case study, as well as a method of locating documents which contain details of cases with a similar pattern to a case being prepared currently. Locating documents which relate to similar matters allows the small operator to use the original documents as precedents or templates for the current case. Often, phrasing, fee structures, outsourcing agencies and methods of solution can be duplicated with very little change to the current matter. To facilitate this process, an intelligent ranking and extraction process can be used to leverage a document search engine and identify matters which relate to the search criteria.

To this end, a number of search requirements were identified in the study:

Name Search - Allows each kind of document stored in the repository to be found by its name. Long file names, spaces and foreign characters are supported by the name search method. Folders can also be retrieved using this method. In the case that the exact name is unknown, a wildcard search would be available.

Full Text Search - This is the most extensive method for conducting a document search. A user is able to search for a single word or phrase, a complex expression, or an extensive range of search standards. Misspellings

can affect the accuracy of the search, which is a major limitation, and this issue needs to be addressed.

Index Search –As the repository becomes large, this method is most suitable. Users can perform alphanumeric character searches to find index information. These searches include wild card expressions and range searches, date and date range searches to locate appropriate matters.

The mercantile agency also maintains electronic notes on matters which are filed in the document repository, the business study identified a requirement to search these notes for pattern matching criteria in order to compare outcomes and legal precedents for dealing with current matters

4.2 Location Of Costs and Fees using the Web

As a means of growing the business, the mercantile agency is looking at accepting the task of debt collection in other localities. In order to process and recover a debt, service fees applicable to the matter may relate to:

- Sheriff costs
- Process Server costs
- Solicitor costs
- Land Titles search fees
- Motor Vehicle records search fees
- ASIC (Australian Securities and Investments Commission) search fees
- Registrar General search fees

All these service providers have sliding fee schedules which require the mercantile agency to determine the fee structure which will apply when dealing with a matter in that particular jurisdiction. Court fees in each jurisdiction may also differ for a particular type of document to be lodged, and these other agencies such as court appointed Sheriffs or investigators who can locate and serve documents on behalf of the mercantile agent must be contacted and paid prior to acting on behalf of the mercantile agency. In addition, court lodgment fees are continually changing, so the source of the fee structure needs to be periodically reviewed to ensure that the current fees are being paid when court documents are lodged. For example, refer to these sites for examples of the disparity in content and style of the fee structure information available in various court localities

http://www.nt.gov.au/ntsc/doc/almanac/2004_almanac.pdf
<http://www.lawlink.nsw.gov.au/sc/sc.nsf/pages/filingfees>
http://www.lawlink.nsw.gov.au/dc.nsf/pages/dc_fees
http://www.courts.tas.gov.au/magistrate/circulars/cir2000_8.htm

Use of the Web to locate information relating to fee comparison, goods and services procurement, as well as access to electoral role information for skip tracing is also an expected area of growth for the mercantile

industry. So the use of intelligent agents to mine the Web for relevant information is of great benefit to all businesses, but particularly relevant to a small business. The major logistical problem for the user of the Web to provide information (at this stage) is the lack of formal content standards as well as the lack of location directories for accessing the data in a structured manner.

As a means of improving office efficiencies, we propose a series of experiments using an ontology driven information system which gathers and builds a knowledge base relating to a specific subject content domain. We intend to build on Brin's Dual Iterative Pattern Relation Extraction (DIPRE) [2] method to construct a semantic ontology of domain patterns which is then used to build the knowledge base. Our system first uses a set of training pages to generate a conceptual map (ontology) of the physical training pages. The application will then use the ontology to search a text repository such as a document management system or the Web. The objective is to rank the results into a meaningful knowledge base of information relating to the subject domain.

5. RELATED WORK

Related work using extraction analysis methods to develop ontologies and information domains by 'trawling' the World Wide Web was begun in earnest at the same time as the Internet became a commercial entity. Research to identify phrase patterns which surround the target domain was begun by Riloff et. al. in 1996[3]. Riloff separated initial content retrieval into relevant and non-relevant domains using a frequency algorithm to determine relevance. Content deemed relevant was then used to generate patterns for subsequent retrieval runs which better filtered the content relating to the target domain. In 1998, Sergey Brin developed a more formal, iterative method for the discovery and containment of patterns (DIPRE - Dual Iterative Pattern Relation Extraction). Brin's approach has been extended by a number of research efforts including Yangarber et al. [4], Soderland and Etzioni (University of Washington) [5] and Agichtein and Gravano (Columbia University) [6] to generate domain ontologies and Q/A query facilities.

Etzioni's group generate ontology repositories using the 'Knowitall' tool they developed. Agichtein's research developed 'Snowball' which also uses an iterative approach using domain seeds to begin the extraction process. Our work is also similar to Crescenzi's Roadrunner [7] project in terms of the use of example pages to build a pattern ontology as a parser for the extracted corpus. Using this method, discovery starts with a few training pages which are known to be good indicators of the topic of interest. These patterns are used to retrieve some relevant content; the content is scanned for associated phrase patterns, and the process

is repeated, refining the target to develop the ontology or specific answer set. In our experiments, we attempt to deliver a tangible efficiency benefit to the user by applying these techniques to the business problem. We use the Web corpus to locate and acquire information on court related fees applicable to the recouping of outstanding debts in court jurisdictions within regional Australia. Our objectives in this paper are to use these techniques to develop semi automated intelligent processes which build patterns from each extraction iteration, then compare these patterns against semantic rules generated from the 'bootstrap' template.

6. INFORMATION RETRIEVAL

6.1 Prelude To Experiments

This section details the architectural design and setup for our experiments. We begin with the development of a browser based request/response interface which interacts with the user, provides input to the Web search engine, Web Service, or application specific search engine. This means the module constructs both a standard html request as well as a SOAP request. An application specific search engine will be expected to accept XML requests, the app server will deliver document URL's to the receiving interface.

When the user begins a session, initialization information is requested, such as an iteration count which limits our run, one or more learning pages which are used to create the ontology and a variable number of key words or phrases which we use to bootstrap our initial search criteria. For our purposes, we define an ontology as a conceptual image of the training pages which we store as a tree of nodes centered around the bootstrap keywords and phrases. We build our tree by first accepting and storing the users search arguments and any example pages used to generate the ontology of patterns including date and time and users id. The contents of the first extraction run are cleaned, to provide input to the pattern maker. The cleaner counts the number of hits of each keyword and passes this to the pattern maker which uses this information as a probability of usefulness. The pattern maker accepts the resultset strings from the receiving interface and clusters [8] them according to occurrence (refer to Figure 2). Patterns are created using the DIPRE algorithm for each pattern identified (keyword or phrase). Each run generates new patterns which are analyzed and compared to the ontology for relevance.

Our initial business goal is to locate and rank pages according to the closeness of the match with the training pages. The system orders the users desired information and builds a knowledge base on URL, location, fee type and fee value.

So on the first run, the discriminators (patterns) which surround the keywords and key phrases are saved and used in the second and subsequent runs. We calculate a probability for the relevance of each page returned and first try to select most obvious discriminators, then loosen constraints and select NEAR discriminators.

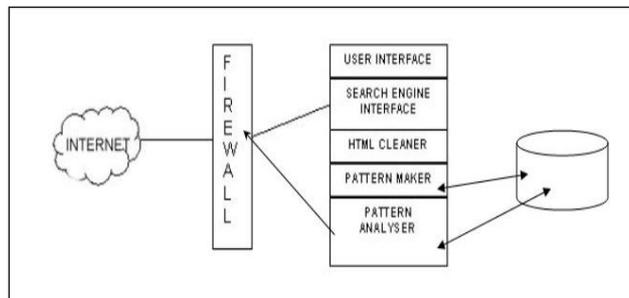


Figure 2.

7. FUTURE WORK

We report the details of our design and results of our experiments in later papers in the series. Our longer term objective is to investigate the introduction of intelligence into Web based applications as a way of improving human/computer interaction efficiencies. We focus our research in the area of small to medium business enterprises (SME's) as this market (SME's) account for more than 90% of Australian businesses [1].

Our work provides us the opportunity to:

- Research and develop applications which utilize the 'semantic' web [9] and related content presentation standards being developed.
- Research the development of agent based software using a genuine commercial environment from which to draw requirements.
- Develop leading edge software solutions which deliver tangible cost saving benefits for small businesses.

REFERENCES

- [1] Collins A.J. et.al.; "Using Portal Technology for Collaborative Ecommerce"; *7th International Conference on Electronic Commerce Research*, Dallas 2004.
- [2] Brin S., "Extracting Patterns and Relations from the World Wide Web", *Proceedings of the 1998 International Workshop on the Web and Databases*, 1998
- [3] Riloff E. "Automatically generating extraction patterns from untagged text". *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 1044-1049, 1996.
- [4] Yangarber R. and Grishman R.; "Extraction Pattern Discovery through Corpus Analysis";

Proceedings of the 2nd International Conference on Language Resources and Evaluation; 2000

[5] <http://www.cs.washington.edu/research/knowitall/>

[6] Agichtein E., Gravano L. ;"Snowball: Extracting Relations from Large Plain-Text Collections"; *Proceedings of the Fifth ACM International Conference on Digital Libraries*; 2000

[7] Crescenzi V. et. al. "The Roadrunner Project: Towards Automatic Extraction of Web Data"; *Universit`a della Basilicata -Italy*; 2001

[8] Bobrovniko D.; "Experiments in Semantic Bootstrapping with a Cluster-Based Extension to DIPRE"; *Stanford University*; 2000

[9] Berners-Lee T. et.al.; The Semantic Web; *Scientific American*; May 2001

Mining the Change of Events in Environmental Scanning for Decision Support

Mengjung Shih¹, Duenren Liu¹, Churnjung Liao², Chinhui Lai¹

¹Institute of Information Management, National Chiao Tung University, Hsinchu, Taiwan, China

²Institute of Information Science, Academia Sinica, Taipei, Taiwan, China

¹{mj_shih, dliu, chlai}@iim.nctu.edu.tw, ²liaucj@iis.sinica.edu.tw

ABSTRACT

An organization's environment is increasingly complex. Business demand on environmental scanning has significantly increased in recent years due to an attempt to assist management in planning an organization's strategies and responses. The conventional technique for environmental scanning is event detection from text documents such as news stories. Event detection methods recognize events while they neglect to discover the changes of events. This work develops an event change detection (ECD) approach that combines association rule mining and change mining techniques. Detecting changes of events aids managers in making fast responses to the change of external environments. Association rule mining is employed to discover the subject behaviors of events from news stories. Changes of events are identified by comparing the subject behaviors of events from different time periods. The discovered event changes can provide effective decision support for decision makers to capture environmental information in a timely manner and make adequate decisions.

Keywords: Environmental Scanning, Change Mining, Association Rule Mining, Event Tracking, Event Detection

1. INTRODUCTION

With the rapid growth of Internet, the business environment today has become more complex and changeable. Environmental scanning has received considerable attention in business in recent years due to its benefits of assisting management in planning the organization's future course of action [4].

News is the best information source for businesses to get environment information (event) because of its comprehensive content. Event detection is the major research field in news, and it plays an important role for supporting environmental scanning [14]. Event detection identifies streams of news stories that discuss new events [1][2][16][17]. Organization managers can be aware of new events in their environment via event detection technologies. Event detection methods are concerned only with recognizing new events and neglect to discover the changes between these news stories. But in fact, changes of environment occur not only through new events but also in existing events.

The purpose of environmental scanning is not only to identify changes of the content of news stories that describe the same or different events, but also to aid business managers to be sensitive to the external environment resulting in a fast response to the environment. Detecting event changes is a critical work for businesses.

An event change is the change of event trends in two time periods. For example, the trend of telecommunications industries in 2001 was to provide educational services to customers, while the trend of telecommunications industries in 2002 was to provide recreational services. The change of service types is named event change.

These trends are discovered from news stories of events- "Services provided in telecommunications industries (telecom services)" in 2001 and 2002 respectively.

To capture the changes of events, we must first determine the subject behaviors of an event. A subject behavior of an event is an incident that is described in most news stories, and it can be characterized by the relationships of 4W properties (When, Who, Where and What). The relationships between 4Ws may change with time, but these changes will be important information for decision making for business managers. For instance, if telecommunication companies were reported to provide recreational services (What property) in most news stories of the event "telecom services", "telecom services" can be viewed as a subject behavior of the event. If subject behaviors of different time periods are different, event change occurs.

In this research, association rule mining is employed to identify the subject behaviors of events from news stories.

Information needs usually vary with time, situation and people. Business managers may require different levels of information in developing different strategies. A decision maker may not only need to know the business operations of individual companies but also the operations of industries. To meet all possible information needs in environmental scanning, it is necessary to construct a concept hierarchy according to the content of news stories.

Motivated by the needs for capturing event changes, the goal of this research is to develop an event change identification (ECI) technique. The proposed technique combines the change mining approach and concept hierarchy to provide useful environment changes to enhance environmental scanning.

The remainder of this paper is organized as follows. Section 2 reviews literatures relevant to this research, including environmental scanning, event tracking and detection, association rule mining and change mining technologies. Section 3 introduces our event changes identification technique. In Section 4, we describe methods to detect event changes. Finally, the contributions of this study are summarized in Section 5.

2. LITERATURE REVIEW

This section reviews literatures relevant to this research, including environmental scanning, event tracking and detection, association rule mining, and changes mining technologies.

2.1. Environmental Scanning

To adapt to the environment and subsequently develop effective responses to secure or improve the business' position in the future, environmental scanning is important [4][7].

Environmental scanning acquires and uses information about a business's external environment, which is used by business managers to make decisions and plans [4]. Consequently, it is essential to scan the external business environment to improve a business's future.

2.2. Event Tracking and Detection

Event tracking and detection techniques are proposed to support people detecting new events and tracking subsequent news stories that discuss previous events. Event tracking starts from a set of pre-classified news stories, and searches out all subsequent stories that discuss the same event [1][17]. The goal of event tracking is to locate follow-up news stories of a manager's event of interest.

Event detection is the major research field in news, and it plays an important role for supporting environmental scanning [14]. The goal of event detection is to identify streams of news stories that discuss new events [1][2][16][17]. Organization managers can become aware of new events in their environment via event detection technologies. But changes of environment occur not only during new events but also during existing events. Business administrators need to be notified before an event takes place, not just when it happens.

2.3. Association Rule Mining

Association rule mining searches for interesting association or correlation relationships among items in a large dataset [6]. There are two measures to represent whether a mined rule meets the criteria of regularity: support and confidence [6][15].

This work proposes an event change identification technique to detect event changes. We apply association

rule mining on news data to detect the regulation of the news.

The format between news data and transaction data is very different. Transaction data is structured, and the scope of attributes and values is often fixed. In contrast to transaction data, news data is unstructured, and the scope of attributes and values often changes. For example, the protagonist of the same event may change.

Therefore, mining association rules from news data needs to address these difficulties. In this study, we used event property selection and extraction and concept hierarchy to solve these problems.

2.4. Change Mining

The objective of change mining is to discover changes of data from different time periods. The researches of change mining can be classified into two main groups:

1. Decision Tree Model: This method first constructs decision trees from different datasets. It obtains differences via comparisons with two decision trees [10]. Through graphic presentation, the user can discover the differentiae and understand the data.

2. Association Rule: This method obtains changes via comparisons with association rules mined from a different dataset [11][12]. The types of changes defined by [5][8][9][12] are emerging patterns (the emerging pattern concept captures significant changes and differences between datasets.), unexpected changes (can be found in newly discovered association rules with consequence/conditional parts different from previous rules), added rules (is a newly appearing rule in the present which could not be found in the past) and perished rules (is a rule that can only be found in the past). Association rule change mining is used to discover changes in customer behavior.

In this study, we adopted the association rule change mining technique to discover changes of events. As mentioned above, transaction data and news data are quite different. We modified the conventional change mining approach, and used event property and concept hierarchy to improve the quality of change detection.

3. EVENT CHANGE DETECTION (ECD) TECHNIQUE

In a dynamic environment, the state of events is constantly changing and evolving. It is important for business managers to capture environmental changes to adjust their business strategies.

3.1. Overview

The proposed event change detection (ECD) technique comprises two main processes: learning and detection. From a set of news stories, the learning process (as shown in Figure 1) seeks to identify event properties, concept hierarchy of property, and association rules representing the subject behaviors of event. The purpose

of the detection process is to identify event change, according to the association rules discovered in the learning process (as illustrated in Figure 2).

3.2. Learning process

The learning process of the ECI technique is to identify which event in the news story should belong and induce event association rules from a set of news stories. The learning process consists of four steps, including news fetcher, event identification, property extraction and association rule matching.

The news fetcher component of the ECD technique is to fetch news stories from online news providers.

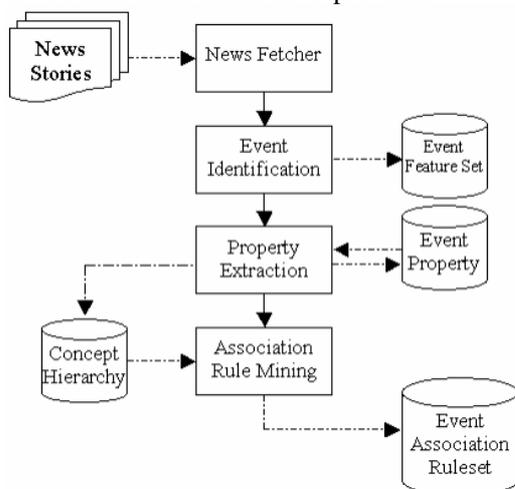


Figure 1. Learning process of identifying event change

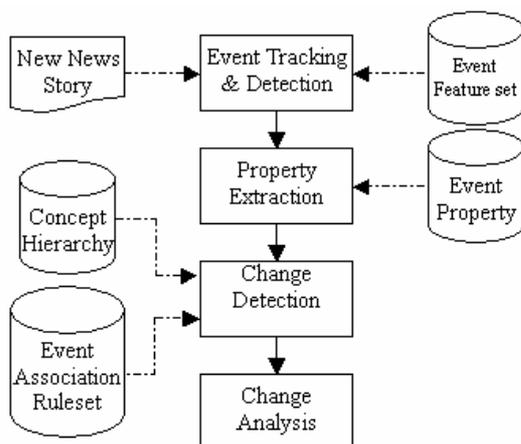


Figure 2. Detection process of identifying event change

3.2.1. Event Identification

The objective of event identification is to determine the set of features that will be used for representing individual news stories within pre-classified training news stories. The search for news categorization patterns uses feature extraction and feature selection. Feature extraction extracts a set of terms based on the news stories. For producing a set of features, we used feature extraction with the Chinese Dictionary [3] which contains

one hundred and sixty thousand terms to parse each training news story to produce a list of terms commonly referred to as features.

Following feature extraction, feature selection is initiated to condense the size of the event feature set (the size of feature set is 30). The purpose of this phase is to remove unnecessary terms from the set that were produced in the previous phase. Several feature selection methods have been proposed in the literature, including TF, TF*IDF and so on. We selected the features in this research by TF*IDF. The top 30 features with the highest feature score are selected as features for representing news stories.

TF*IDF is defined as [6]:

$$TF \times IDF(t, d) = tf(t, d) \times \log(N/n_t) \quad (1)$$

where

N is the size of the news stories in the collection;

n_t is the number of news stories where t occurs;

$tf(t, d)$ is the term frequency (TF) within a news story d ;

$\log(N/n_t)$ is the inverted document frequency (IDF).

3.2.2. Event property

Feature-based technique is one of the most popular methods to represent the content of a news story. But feature-based representation often performs poorly because of its inherent problems, such as vocabulary discrepancies between reporters, news stories for different events containing similar feature sets, and so on. To overcome the problems of the traditional feature-based technique, event property has been used to improve event detection [1][14]. To improve the accuracy of content representation, we adopted event property to identify the subject behaviors of events. Event properties are classified into four categories (4Ws) as summarized below.

1. When: Date, Time
2. Who: Person, Organization
3. Where: Location
4. What: Action, Claim, Standpoint, Statement

3.2.3. Concept hierarchy

As previously mentioned, business decision makers may require different levels of information for developing different strategies. Such different levels of information form a concept hierarchy. In this work, we adopted concept hierarchy to meet all possible information needs. Concept hierarchy is used to define the relations of concept levels between a group of attributes [13].

In the process of event change detection, we need to determine the difference between the values of attributes. Subject behaviors of two time periods are compared. These values can be regarded as the nodes of concept hierarchy.

The node difference in a concept hierarchy is:

$$h_H(A, B) = \frac{\text{Max}\left(\sum_{L_i \in P_A} WL_i, \sum_{L_j \in P_B} WL_j\right) - \sum_{L_k \in P_{\text{comm}}(A, B)} WL_k}{\text{Max}\left(\sum_{L_i \in P_A} WL_i, \sum_{L_j \in P_B} WL_j\right)} \quad (2)$$

A and B are nodes in concept hierarchy
 P_A : the path from root to node A.
 P_B : the path from root to node B.
 P_{comm(A, B)}: the common path between P_A and P_B.
 L_i: a link in P_A.
 L_j: a link in P_B.
 WL_i: the weight on the level of link i.
 WL_j: the weight on the level of link j.

3.3. Event tracking and detection

The event tracking and detection process identifies which event a recent news story belongs to and whether the news story discusses a new event. To achieve this, the event tracking and detection process consists of three steps: news document representation, similarity comparison and event assignment.

News document representation:

Each news document is represented using representative features. The feature set for each event is extracted and selected from a set of labeled training news stories.

Similarity comparison:

This step is to calculate the similarity between recent news story D and all known events. The cosine distance is used to compute the similarity, as shown as below:

$$\text{Sim}(D, e) = \frac{\sum_{f_k \in D \cup e} t_{Dk} \times t_{ek}}{\sqrt{\sum_{f_k \in F_D} t_{Dk}^2} \sqrt{\sum_{f_k \in F_e} t_{ek}^2}} \quad (3)$$

where F_D and F_e are the feature sets of document D and event e respectively. t_{Dk} is the weight of the representative feature f_k of D, and t_{ek} is the weight of the representative feature f_k of e.

Event assignment:

Upon obtaining the similarity scores between the latest news story D and all known events, we set a pre-specified threshold in this step, and assign an event label according to similarity scores. If the maximum similarity score between D and the known events is below the threshold, the new document is labeled as the first story of a novel event; otherwise, we assign the news document to the event where the maximum similarity score is obtained.

4. DETECTION OF EVENT CHANGE

Event change is the change of event trends in two time periods. Those trends are discovered from news stories of the same event during different times. In order to detect event changes from a great quantity of news stories, we extended the idea of change detection methodology [13] and integrated concept hierarchy into the event change detection technique. The detection of event change is revealed in Figure 3.

Based on past research and business requirements, there are five types of possible mining changes: emerging

patterns, unexpected consequent changes, unexpected condition changes, added rules, and perished rules.

4.1. Discovery of changed rule

In order to discover the five types of event changes between different time periods, we employ the rule matching method proposed by Song et al. [12] and modify it to fit event change detection by considering concept hierarchy.

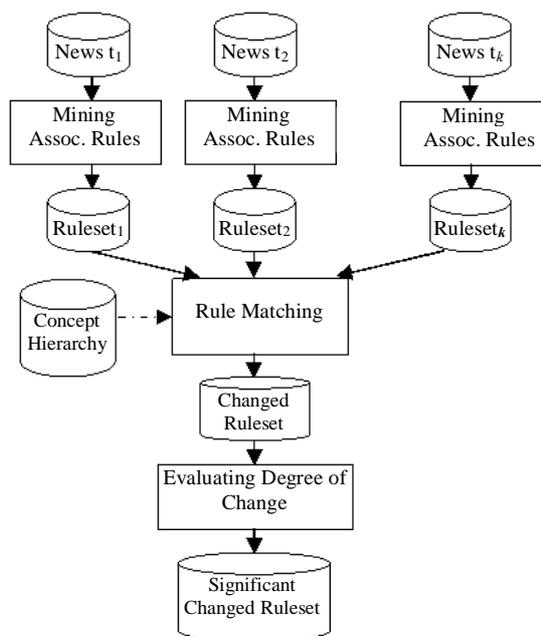


Figure 3. The process of event change detection

The rule matching method consists of three steps:

1. Calculate the maximum similarity value. For each rule in different time periods, calculate the maximum similarity value.
2. Calculate the difference measure. For each rule r_i^t in time t, calculate the difference measures between r_i^t, r_j^{t+k}.
3. Decide type of changes. According to the judged factors (maximum similarity value and the difference measures), this step determines the type of change for the rules.

Before we introduce the methods to calculate judged factors, some notations are defined as below.

Conditional part of rules:

p_{ij} : Degree of attribute match of conditional part

$$p_{ij} = \frac{|A_{ij}|}{\max(|X_i^t|, |X_j^{t+k}|)}$$

A_{ij} : Set of attributes common to both conditional parts of r_i^t and r_j^{t+k}

X_i^t : Set of attributes in the conditional parts of r_i^t

X_j^{t+k} : Set of attributes in the conditional parts of r_j^{t+k}

A_{ijk} : The k^{th} attribute in A_{ij} .
 $v(r_i^t, A_{ijk})$: Value of the k^{th} attribute in A_{ij} of r_i^t .
 $v(r_i^{t+k}, A_{ijk})$: Value of the k^{th} attribute in A_{ij} of r_i^{t+k} .
 l_{ijk} : Degree of value match of the k^{th} matching attribute in A_{ij}
 $l_{ijk} = 1 - h_H(v(r_i^t, A_{ijk}), v(r_i^{t+k}, A_{ijk}))$

Consequent part of rules:

q_{ij} : Degree of attribute match of consequent part
 $q_{ij} = \frac{|B_{ij}|}{\max(|Y_i^t|, |Y_i^{t+k}|)}$
 B_{ij} : Set of attributes common to both consequent parts of r_i^t and r_i^{t+k}
 Y_i^t : Set of attributes in the consequent parts of r_i^t
 Y_i^{t+k} : Set of attributes in the consequent parts of r_i^{t+k}
 B_{ijm} : The m^{th} attribute in B_{ij} .
 $v(r_i^t, B_{ijm})$: Value of the m^{th} attribute in B_{ij} of r_i^t .
 $v(r_i^{t+k}, B_{ijm})$: Value of the m^{th} attribute in B_{ij} of r_i^{t+k} .
 f_{ijm} : Degree of value match of the m^{th} matching attribute in B_{ij}
 $f_{ijm} = 1 - h_H(v(r_i^t, B_{ijm}), v(r_i^{t+k}, B_{ijm}))$

The similarity measure S_{ij} between r_i^t and r_i^{t+k} is calculated using the following formula: ($0 \leq S_{ij} \leq 1$)

$$S_{ij} = \begin{cases} \frac{p_{ij} \times \sum_{k=1}^{|A_{ij}|} l_{ijk}}{|A_{ij}|} \times \frac{q_{ij} \times \sum_{m=1}^{|B_{ij}|} f_{ijm}}{|B_{ij}|}, & \text{if } |A_{ij}| \neq 0 \text{ and } |B_{ij}| \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Where $\frac{p_{ij} \times \sum_{k=1}^{|A_{ij}|} l_{ijk}}{|A_{ij}|}$ represents a similarity of conditional part, and $\frac{q_{ij} \times \sum_{m=1}^{|B_{ij}|} f_{ijm}}{|B_{ij}|}$ represents a similarity of consequent part between r_i^t and r_i^{t+k} . If the conditional parts and consequence parts r_i and r_i^{t+k} are the same, S_{ij} equals 1.

Accordingly, the maximum similarity value of r_i^t is as below: $V_i = \max(S_{i1}, S_{i2}, \dots, S_{i|R_i^t|})$

Accordingly, the maximum similarity value of r_i^{t+k} is as below: $V_i = \max(S_{i1}, S_{i2}, \dots, S_{i|R_i^{t+k}|})$

The second judged factor, difference measure ∂_{ij} between r_i^t and r_i^{t+k} is given by ($-1 \leq \partial_{ij} \leq 1, |\partial_{ij}| \leq 1$)

$$\partial_{ij} = \begin{cases} \frac{p_{ij} \times \sum_{k=1}^{|A_{ij}|} l_{ijk}}{|A_{ij}|}, & \text{if } |A_{ij}| \neq 0 \text{ and } |B_{ij}| = 0 \\ \frac{p_{ij} \times \sum_{k=1}^{|A_{ij}|} l_{ijk}}{|A_{ij}|} - \frac{q_{ij} \times \sum_{m=1}^{|B_{ij}|} f_{ijm}}{|B_{ij}|}, & \text{if } |A_{ij}| \neq 0 \text{ and } |B_{ij}| \neq 0 \\ -\frac{q_{ij} \times \sum_{m=1}^{|B_{ij}|} f_{ijm}}{|B_{ij}|}, & \text{if } |A_{ij}| = 0 \text{ and } |B_{ij}| \neq 0 \end{cases} \quad (5)$$

If attributes of conditional parts and consequent parts between the two rules are different, it makes no sense to compare the degree of difference, i.e., $|A_{ij}| = 0$ and $|B_{ij}| = 0$, because these two rules are completely different.

Although r_i^{t+k} is determined to be an unexpected change with regard to r_i^t by ∂_{ij} , it may be an emerging pattern of other rule. An emerging pattern may be misjudged to be an unexpected change. To improve the accuracy of judging change type, the third judged factor, modified difference measure ∂'_{ij} is denoted as:

$$\partial'_{ij} = |\partial_{ij}| - j_{ij}, \text{ where } j_{ij} = \begin{cases} 1, & \text{if } \max(V_i, V_j) \geq q \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

The parameters q_1, q_2 , and q_3 , (event change thresholds) are thresholds to determine the types of change. The five types of change can be classified according to the three kinds of judged factors and the three predefined thresholds: q_1 for emerging patterns, q_2 for unexpected consequent and unexpected condition changes, and q_3 for added and perished rules. In general, $q_1 > q_2 > q_3$, and q_1 approaches to 1. Table 1 shows the value of measure for each type of change.

Table 1. Value of measure for each type of change

Type of Change	Value of measure to classify
Emerging Pattern	$S_{ij} > q_1$
Unexpected Consequent	$\partial_{ij} > 0, \partial'_{ij} \geq q_2$
Unexpected Condition	$\partial_{ij} < 0, \partial'_{ij} \geq q_2$
Added Rule	$V_j < q_3$
Perished Rule	$V_i < q_3$

4.2. Evaluating degree of event changes

Changes of environment are varied. Managers need to focus on essential changes. To achieve this goal, it is important to evaluate the degree of change, and rank changed rule by degree. Table 2 shows a simple formulation of change degree evaluation.

Table 2. Degree of event change

Type of Change	Degree of Change
Emerging Pattern	$\frac{Support^{t+k}(r_i) - Support^t(r_i)}{Support^t(r_i)}$

Unexpected Change	$\frac{Support^t(r_i) - Support^{t+k}(r_i)}{Support^t(r_i)} \times Support^{t+k}(r_i)$
Perished Rule	$(1-V_i) \times Support^t(r_i)$
Added Rule	$(1-V_i) \times Support^{t+k}(r_i)$

Let $support^t(r_i)$ and $support^{t+k}(r_i)$ represent the support value of r at time t and $t+k$, respectively. Degree of change for emerging pattern shows the change of support value between time t and time $t+k$. Degree of change for unexpected change is the change ratio of r_i multiplied by the support value of r_i at time $t+k$. The change ratio of r_i represents the degree of unexpectedness of r_i , namely the proportion of the difference between $support^t(r_i)$ and $support^{t+k}(r_i)$ to the $support^t(r_i)$. The degree of change is larger, if the change ratio of r_i is larger and $support^{t+k}(r_i)$ is larger. Degree of change for perished (added) rule is obtained from the support value of perished (added) rule multiplied by the value of 1 minus the maximum similarity value. The degree of change is larger, if the perished (added) rule has less maximum similarity value.

5. CONCLUSION

The business environment today is more and more complicated. The ability to capture environment changes and be sensitive to the environment becomes a critical success factor for business. Research on environment scanning currently puts great emphasis on event tracking and event detection. But the main purpose of event tracking and event detection is to identify which event the news story describes. Business administrators cannot obtain environment change information until a new event occurs. It is insufficient for managers to be notified only when a new event happens.

To capture the changes of events, this research proposes an event change detection (ECD) technique. The proposed technique combines the change mining approach and concept hierarchy to detect changes to enhance environmental scanning.

REFERENCES

- [1] Allen J., Papka R. and Lavrenko V., "On-line new event detection and tracking," in *Proc. of 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM Press, Melbourne, Australia, pp.37-45, 1998.
- [2] Brants T., Chen F. and Farahat A., "A system for event detection," in *Proc. of 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM Press, Toronto, Canada, 2003, pp.330-337.
- [3] Chinese Dictionary, <http://william.cswiz.org/techreport/moecdicit>
- [4] Choo C. W., "The art of scanning the environment," *Bulletin of the American Society for information Science and Technology*, vol.25, no.3, pp.21-24, 1999.
- [5] Dong G. and Li J., "Efficient mining of emerging patterns: discovering trends and differences," in *Proc. of 5th International Conference on Knowledge Discovery and Data Mining, KDD-99*, 1999, pp.43-52.
- [6] Han J. and Kamber M., *Data Mining-Concepts and Techniques*, Morgan Kaufmann Publishers, San Francisco, 2001.
- [7] Jennings D. and Lumpkin J., "Insights between environmental scanning activities and porter's generic strategies: an empirical analysis," *Journal of Management*, vol.18, no.4, pp. 791-803, 1992
- [8] Lanquillon C., "Information filtering in changing domains," in *Proc. of the International Joint Conference on Artificial Intelligence, IJCA99*, pp.41-48, 1999.
- [9] Liu B. and Hsu W., "Post-analysis of learned rules," in *Proc. of 13th National Conference on Artificial Intelligence, AAAI-96*, pp.828-834, 1996.
- [10] Liu B., Hsu W., Han H.-S. and Xia Y., "Mining Changes for Real-life Applications," *the Second International Conference on Data Warehousing and Knowledge Discovery (DaWak)*, pp.337-346, 2000.
- [11] Liu B., Hau W. and Ma Y., "Discovering the Set of Fundamental Rule Changes," in *Proc. of the 7th ACM International Conference on Knowledge Discovery and Data Mining*, 2001.
- [12] Song H. S., Kim J. K. and Kim S. H., "Mining the change of customer behavior in an internet shopping mall," *Expert Systems with Applications*, vol. 21, pp.157-168, 2001.
- [13] Srikant R. and Agrawal R., "Mining Generalized Association Rules," in *Proc. of the 21th VLDB Conference*, Zurich, Switzerland, 1995.
- [14] Wei C.-P. and Lee Y.-H. "Event detection from online news documents for supporting environmental scanning," *Decision Support Systems*, vol.36, pp. 385-401, 2004.
- [15] Witten I. H. and Frank E., "Output: Knowledge Representation," in *Data Mining*, Morgan Kaufmann Publishers, San Francisco, 2000.
- [16] Yang Y., Pierce T., and Carbonell J. G., "A study on retrospective and on-line event detection," in *Proc. of 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM Press, Melbourne, Australia, 1998, pp.28-36.
- [17] Yang Y., Carbonell J. G., R. D. Brown, T. Pierce, B. T. Archibald and X. Liu, "Learning approaches for detecting and tracking news events," *IEEE Intelligent Systems*, vol.14, pp.32-43, 1999.

Monitoring the Supply of Products in a Supply Chain Environment: a Fuzzy Neural Approach

Chunling Liu¹, Bo Meng¹, Jizi Li²

¹Computer School, Wuhan University, Wuhan 430072, China

²school of Economics & Management, Wuhan University of Science & Technology, Wuhan 430073, China
liuchunring@yahoo.com

ABSTRACT

The fuzzy logic is applied to resolve the monitoring problem of products quality and products quantity increasingly varying as market requirement. A series of fuzzy rules are employed and the fuzzy system may generate suggested supply change rate. At the same time, the operation of supplier is also dynamically changing and the evaluation and selection for supplier are the basis of supply-chain co-operation. So whether it is scientific to select a supplier is crucial for sustaining and developing a company. Therefore, in this paper the neural network is introduced to dynamically assess suppliers and recommend to substituting for new ones when necessary, only supplementing fuzzy logic system with its advantages. This paper describes the methodology for the deployment of this proposed hybrid approach to enhance the machine intelligence of a supply chain network with the description of a case study to exemplify its underlying principles.

Keywords: supply chain, product monitoring, fuzzy logic, neural network

1. INTRODUCTION

It has been found by statistics that the top criteria of enterprises selecting suppliers are product quality. In other words, how the output products quality be is vital for a company's sustaining. With the variation of required output products quality and quantity, it is essential to decide the supply quality and quantity. Thereby, monitoring the supply products quantity and quality in a supply chain is a crucial factor for company to win competency and consolidate the supply chain.

An efficient supply chain network needs to have the capability of assessing the performance among various business partners, thereby enabling the effective selection of enterprise members to form an alliance for certain value chain activities such as the provision of certain products to a certain marketplace. In this respect, every partner contributes its core capability to the alliance in its certain dominant field. So the alliance members may make up for the demerits commutatively to cope with the increasingly complex setting of organizations and markets.

As such a 'smart' approach which allows a company to assess and to select business partners to jointly undertake a business project with the focus on customer demands is critical to leverage the operational performance of the network. To tackle the issue, a product monitoring approach is discussed in details in this paper and the monitoring problem for required products quality and quantity is analyzed based on the real time market, and meanwhile developed the performance assessment and selection suggestion of business partners focusing on fuzzy logic principle and neural network training. In short, the hybrid scheme which takes full advantages of the two reciprocal

intelligence technologies, will support the whole supply chain network to compute the updated performance scores of candidate supplier partners, hereby comprehensively evaluate and select supplier partners and, subsequently, recommend of the exact quantity of supply goods for various partners. In order to further the proposed approach, an example is cited to fully demonstrate the feasibility of implementing the suggested methodology.

2. ASSOCIATED RESEARCH

Fuzzy logic is a problem-solving control system methodology that lends itself to implementation in systems ranging from simple, small, embedded micro-controllers to large, networked, multi-channel PCs of workstation-based data acquisition and control systems (Kaehler, 1998). In other words, it is a method of reasoning that allows a fuzzy description of very flexible rules. But it must be sure that fuzzy logic does not mean fuzzy answers for only the reasoning process is set on vague terms. In fact, many objects and method to need decision and choice in the real world are hard to describe by discrete mathematic module. Fuzzy rules probed a methodology for dealing with the gray areas that involve fuzzy concepts. By allowing rules to have blurry boundaries rather than sharp ones, rules are more flexible, compact and intuitive for modeling complicated processes whose components cannot be clearly separated into categories. Fuzzy logic system involves information fuzzification, fuzzy reasoning and defuzzification. Therefore, as stated above, fuzzy logic does not mean fuzzy answers. Indeed, a discrete answer will be generated through the process of data fuzzifying, fuzzy reasoning and crisp computation.

In general, the knowledge exists plainly in a fuzzy-logic

system that has the advantage of compact, intuitive in algorithm and low in data quantity of processing. However, fuzzy reasoning sacrifices some explanation for accuracy, reliability and compactness, and often cannot explain how outputs change in response to inputs. Fuzzy rules are usually established on field experts and the basis of factual experiments. However, these experts may not be able to state their decision-making process in detail and it may not be feasible and economical for the experts to be around all the time.

In contrast with fuzzy logic, neural networks can be used with data promising results where there are data, even without experts to make judgments about the data. They were found to be very useful at modeling large amounts of data that help to solve complex problems. Indeed, neural networks can be powerful for problem solving because of their predict-compare-adjust process by which the result is next to optimum. In brief, neural network systems have great self-adjusting and self-studying ability and full parallel character in structure. However, knowledge is stored at net nodes in the course of NN processing, therefore, NN system is difficult to describe itself specifically, large in quantity of information processing and much complicated in structure.

As stated above, the two intelligent technologies both have merits and defects. So a hybrid approach which focuses on the fuzzy logic scheme with the addition of the neural network technology will enable the two to complement and supplement each other with their own advantages. However, this research domain in which the synthetic approach is applied to support of a complex supply chain network has not received the attention it deserves. This paper attempts to provide some insights for the framework with a demonstration to show its feasibility.

3. THE FUZZY LOGIC APPROACH OF MONITORING PRODUCTS SUPPLY

3.1 Adapted background

The paper is focused on procurement-quantity allocation policy of goods supplied for distribution center prior to selling to the public. But the scheme proposed is also adapted for the problem of procurement-quantity allocation among suppliers in the process of products procurement of enterprises, given the total required quantity and sure quality.

Products quality is variable with the different suppliers in a complex supply chain environment and the different allocation of products supply quantities among them will lead to the fluctuation of the output products quality. In real supply situation, the overall products quantity and quality required in the distribution center are ever-changing in corresponding with real-time market dynamically. Since the change rate of these data is often

vague, the products quantity allocated for suppliers is also blurry. Therefore, fuzzy logic system applied for this complex supply situation may provide the validity of decision-problem-solving. In the following, the paper offers a simplified scenario of actual circumstances that only involves two suppliers. This case demonstrates that the fuzzy logic approach can deal with this situation and the same approach can be applied to multiple suppliers.

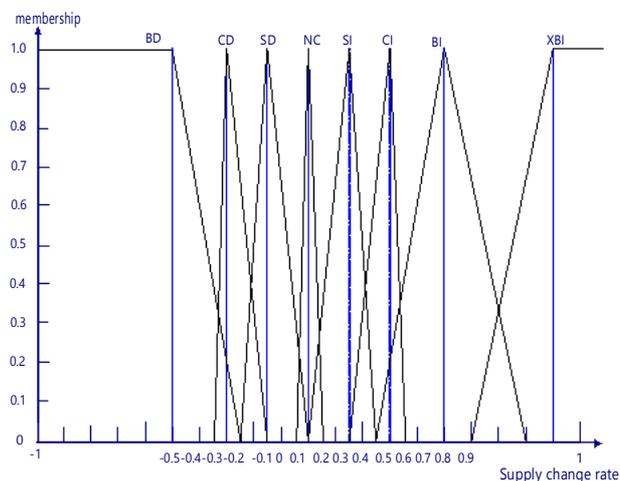


Figure 1: fuzzy pattern of supply change

3.2 Establishment of fuzzy rules

Product quality is variable with the different suppliers in a complex supply chain environment and the different allocation of product supply quantities among them will lead to the fluctuation of the output product quality. On the other hand, the product supply quantities of suppliers vary with every of the two criteria that mean product quality and quantity required by enterprises. In order to provide suggestion supply change rate, a series of fuzzy rules are created to monitor the overall quality and quantity of the supplying products. The fore wares and post wares of rules both encompass fuzzy terms. For example,

If (the overall product quality can be considerably improved and the overall quantity can be significantly increased) Then (Maintain the supply of company-B and increase the supply of company-A significantly)

The meaning of the above fuzzy rule is plain. As the product quality can be considerably improved, i.e. the product quality is considerably lower than the market can appreciate, it is acceptable to supply more units from company A (higher quality supplier), and maintain the quantity of supplier B (lower quality supplier), thereby increasing the overall product quantity at the same time.

Output quality	Output quantity	Supplier A	Supplier B	Output quality	Output quantity	Supplier A	Supplier B
I	L	XBI	NC	OK	CH	SD	SD
I	CL	BI	SD	OK	H	CD	CD
I	OK	BI	CD	CS	L	NC	BI
I	CH	CI	CD	CS	CL	NC	CI
I	H	CI	BD	CS	OK	SD	SI
CI	L	CI	SI	CS	CH	SD	NC
CI	CL	CI	NC	CS	H	CD	NC
CI	OK	CI	SD	S	L	SD	XBI
CI	CH	SI	SD	S	CL	CD	XBI
CI	H	SD	CD	S	OK	CD	BI
OK	L	CI	CI	S	CH	CD	CI
OK	CL	SI	SI	S	H	BD	CI

Table 1: Control rules of supply change rate

The fuzzy rules that the actual output products quality and quantity vary as market requirement can be found in Table 1. One of the fore wares, the output product quality is divided into five degrees in fuzzy pattern which are inferior (I), considerably inferior(CI), ok(OK), considerably superior(CS), superior(S). The other fore wares, the output product quantity is also divided into five degrees in fuzzy pattern which are low (L), considerably low (CL), ok (OK), considerably high(CH), high(H). Additionally, the supply change rate of suppliers is fuzzified as 8 degrees that are extendedly-big increase (XBI), big increase (BI), considerably increase (CI), small increase (SI), no

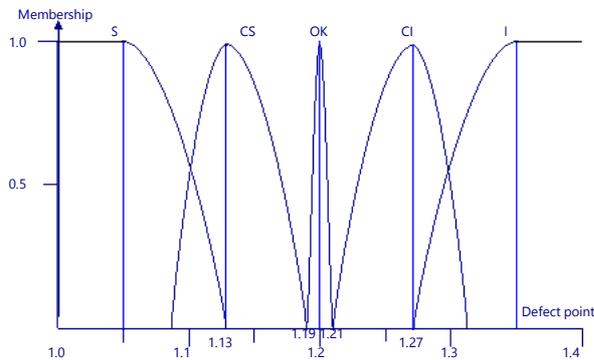


Figure 2 : Fuzzy set for output quality

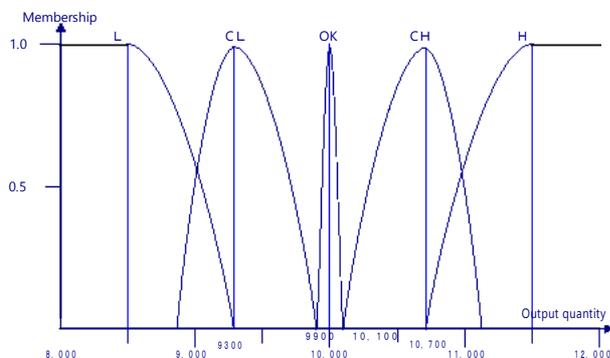


Figure 3: Fuzzy set for output quantity

change (NC), small decrease(SD), considerably

decrease(CD), big decrease(BD). The fuzzy rules based on Table 1 may be created in fuzzy-CLIPS, For example,

```
(defrule quality-inferior-quantity-considerable-low)
(output-quality inferior)
(output-quantity considerable low)
=>
(supplier-A BI)
(supplier-B SD)
```

Figure 1 shows the fuzzy pattern of supply change rate. It is stated that these data are determined on the basis of field knowledge and past experience.

3.3 Fuzzification and fuzzy reasoning

It is assumed that products quality of supplier A and B is examined by defect point (DP). And the DP of supplier A and B is separately 0.8 and 1.5. Any single quality-problem such as color mismatch of a certain component or malfunction of a certain feature can be regarded as a DP. It is key for seeking optimum allocation rate of supply goods among suppliers, when it is targeted for reaching the best price and maintaining quality level as high as possible. In this case it is promised that the integrated quality level is 1.2DP, and the promising output quantity is 10000 per day. The current supply quantity of supplier A is 4150 per day, while that of supplier B is 3850 per day. Then the overall output quality (Q₀) of the combined product is $(4150*0.8+3850*1.5)/(4150+3850)=1.14DP$.

That is to say the actual output quality is 1.14DP while the required quality output is 1.2DP. Referring to Figure 2, the measured value should be categorized as considerably superior quality through fuzzified analysis by the fuzzy expert systems. In the same way, the current overall products quantity is 8000 per day, belonging to low quantity (L) in fuzzy sets (Figure 3).

The conclusion through reason by fuzzy rules (Table 1) after fuzzifying the actual output figures is maintaining the supply quantity of supplier A (NC), while significantly increasing that of supplier B(BI).

3.4 Defuzzified process

Fuzzy logic is a process that involves data fuzzification, fuzzy reasoning, fuzzy decision and, clarifying linguistic values. In order to achieve a crisp value for the problem, the next step is to defuzzify the results generated from the fuzzy interference process by applying a single discrete action. Crisping computation is also known as defuzzification process referring to transforming a fuzzy set to a crisp value.

Defuzzification can be done with several methodologies, such as mean of maxima, center of area and mean of increasing gravity. As for all methods, there are advantages and drawbacks with each. However,

whichever of those you choose, special attention and consideration should be paid to the company's constraints on the system. The details of each method can be found easily on some references and will not be discussed again here. Referring to Figure 1, if the mean of maxima is considered, the supply change rate of supplier B is 0.5(BI in fuzzy set), while that of supplier A is no change (NC in fuzzy set). Therefore, the new supply quantity of supplier B is $3850+3850*1.5=5775$

3.5 Evaluating the final results

The new overall products quantity after adjusting the supply of supplier B is

$$4150+5775=9925$$

The new overall products quality is

$$(4150*0.8DP+5775*1.5DP)/9925=1.206DP$$

From Figure 2 and Figure 3, it is not seen that the new output results are both within the OK level. The fuzzy sets are concretely classified in the paper with a view to achieving propel output result after one fuzzy resolving procedure. However, if the new supply quality is out of the OK range, another mending rule will be fired to obtain the appropriate output values.

4. ASSESSING SUPPLIER DYNAMICALLY WITH NN NETWORKS

4.1 modeling

Artificially neural network (NN) is a complex net system broadly interconnected by a great deal of simple processing cells (neural cells). The neural network which mimics many foundational functionalities of human brain is also a highly complicated nonlinear dynamic system. NN is especially adapted to deal with indiscrete and linguistic information-processing problem required for giving a thought to great many of factors and conditions at the same time, because of its abilities of large-scale parallel, distributed storing and processing, self-organizing, self-adjusting and self-studying.

However, as to the monitoring for product supply, it is far from sufficiency to only pay attention to the supply distribution among suppliers. Because the performance of suppliers is also an ever-changing factor and the evaluation and selection for suppliers is the operating basis of supply chain correlation. In this paper, a neural network module is introduced to assess the suppliers dynamically and suggest the replacement of suppliers when necessary, only supplementing the fuzzy logic system.

The comprehensive performance of suppliers is constrained by multiple factors such as price, in-time delivery, quality, technologic potential, service after selling, historic performance and etc. It is a non-structured information-searching problem to apply

NN to the evaluation of suppliers' performance. It is practiced that application of NN in this respect is effective. A NN module includes two parts: network training and operation evaluation of suppliers. First of all, quantities of historic data required no less than 100 data sets should be collected around suppliers' performance and potential business problems. The next step is training the net on the basis of BP neural network. A reliable automatically operating network may be generated after repeated predict-compare-adjust process. A neural network consists of one input layer, one or more hidden layers, and one output layer. Connections exist between the nodes of adjacent layers to relay the output signals from one layer to the next. Once the net gravities are trained well, the neural network may be used to evaluate the operation and provide decision support for choosing promising suppliers.

As Figure 4 shows, the input layer in the module encompasses four characteristic values (that are price, quality, delivery efficiency, service after sell), each of which selects five recent related records. The output layer is the five assessment suggestions as for suppliers (that are Price decreases slightly, Suggestion to replace supplier, Further assessment of company, Quality

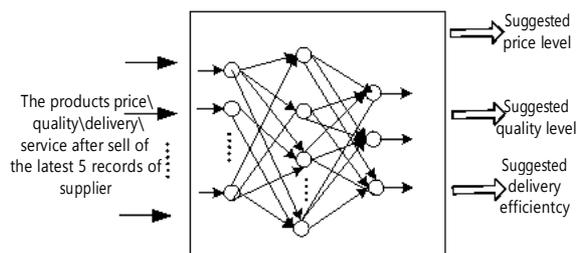


Figure 4: The neural network model of partner evaluation

increases slightly, Delivery time seems to be inconsistent). The database has acquired a great deal of useful data. Accordingly, NN model may create a reliable well-trained network.

4.2 The analysis of a case

Here the further analysis about the reliably trained neural network from a case indicates the feasibility of the neural network system for evaluation and selection of suppliers. At first, the input data should be transformed as one scale. In concrete the five latest track records of a company gathered are scored in performance according to the promising price, quality, delivery efficiency and service level after sell for goods of suppliers. The performance score point (PSP) ranges from 1(lowest score) to 10(highest score). The PSP values are the input of the neural network module which will generate an assessment report for the user to make decisions. The case of Table 2 shows the recent PSP of

supplier A. The output report (Table 3) may be obtained which can be referred to making decision, after inputting the PSP values into the NN modules. In Table 3, “1” output from the NN mode indicates a positive suggestion to the associated statement and “0” is the negative suggestion, whereas “0” means that the related parameter almost is no change.

Table 2:PSPof supplier A

Supplier A(latest record)	Product price PSP	product quality PSP	Delivery schedule PSP	Service after sell PSP
1st	7.5	6.0	8.0	8.0
2nd	7.5	6.5	7.4	7.5
3rd	7.5	7.0	7.5	7.7
4th	7.5	6.0	7.8	7.5
5th	7.5	4.5	7.5	7.8

Table 3:Assessment report of supplier A

Supplier A	NN output
Price decrease slightly	0.5
Replacement of supplier	0
Quality is increasing slightly	0
Need to further assessment	1
Inconsistent delivery efficiency	0.5

As for the report of Table 3, it is obvious that there is question about supply quality of supplier A. In general, if the supplier A is one of main suppliers of the company, the further investigation of the issue should be put on. And if the investigation shows that supplier A has tackled the current difficulty through some urgent steps. Then the evaluation result is determined. Hereby, users can easily and objectively evaluate collaborative partners, find out problem timely and seek business chances without specialists or experts' instructions, according to the output suggestion of the NN model..

5. CONCLUSION

The problems are analyzed in the paper how enterprises assess and select suppliers and how to decide the propel allocation among multiple suppliers about procurement quantity in complex supply chain environment so as to meet the requirement of the dynamical overall products quality. The fuzzy logic is applied to resolve the monitoring problem of products quality and products quantity increasingly varying as market requirement. A

series of fuzzy rules are fired and the fuzzy system may generate suggested supply change rate. In the mean while, the operation of supplier is also dynamically changing and the evaluation and selection of suppliers are the basis of supply –chain co-operation. So whether it is scientific to select a supplier needs to be emphasized for sustaining and developing a company. Therefore, in the paper the neural network is introduced to dynamically assess suppliers and suggest to substituting for new ones when necessary, only supplementing fuzzy logic system with its advantages.

However, the hybrid approach in the paper is by no meant perfect. Fuzzy logic needs generating rules by experts while NN needs an array of historic data to train a network complexly. However, the paper is aimed at constructing an intelligently technologic flat for reference in the complicated supply chain management, aiding decision-making, thus making SCM more “agile” and enhancing management efficiency.

ACKNOWLEDGEMENT

We are most grateful to the School of computer, Wuhan University and professor ZENG Guang-qi of school of Electronics & Communication, Wuhan University of science & technology.

REFERENCES

- [1] DHAR V. and H. HELLENDORRN, and M.REINFRANK(1996),An Introduction to Fuzzy Control,Heidelberg:Springer,149~163
- [2] MARK BRACKSTONE, “Examination of the use of fuzzy sets to describe relative speed perception”, ERGONOMICS,VOL.43, NO.4, P528~542,2000
- [3] H.C.W.Lau,I.K.Hui,etc. “Monitoring the supply of products in a supply chain environment: a fuzzy neural approach”, Expert Systems,Vol.19, No.4,P235~243,2002
- [4] WH Wenbin, “Research on the fuzzy neural network in the decision support system”, Systems Engineering and Electronics, Vol.25, No.12,2003
- [5] Zadeh A, “Fuzzy sets and fuzzy information granulation theory”, Beijing: Beijing Normal University Press, P121~183,2000

PATSEEK: Content Based Image Retrieval System for Patent Database

Avinash Tiwari¹, Veena Bansal²

^{1,2} Industrial and Management Engineering Department, IIT Kanpur 208016 India
²veena@iitk.ac.in

ABSTRACT

A patent always contains some images along with the text. Many text based systems have been developed to search the patent database. In this paper, we describe PATSEEK that is an image based search system for US patent database. The objective is to let the user check the similarity of his query image with the images that exist in US patents. The user can specify a set of key words that must exist in the text of the patents whose images will be searched for similarity. PATSEEK automatically grabs images from the US patent database on the request of the user and represents them through an edge orientation autocorrelogram. L1 and L2 distance measures are used to compute the distance between the images. A recall rate of 100% for 61% of query images and an average 32% recall rate for rest of the images has been observed.

Keywords: Patent Search, Content Based Image Retrieval, Recall rate, Precision

1. INTRODUCTION

The sheer size of the information available about the patents has led many researchers to develop efficient and effective retrieval techniques for patent databases. The Web Patent Full-Text Database (PatFT) of United States Patent Office (USPTO) contains the full-text of over 3,000,000 patents from 1976 to the present and it provides links to the Web Patent Full-Page Images Database (PatImg), which contains over 70,000,000 images. The volume of this repository makes it prohibitive for humans to find similar images in them. This has motivated us to develop a content-based image retrieval system for the patent databases.

Initial image retrieval techniques were text-based that associated textual information, like filename, captions and keywords with every image in the repository. For image retrieval, keyword based matching was employed for finding the relevant images. The manual annotation required prohibitive amount of labor. Moreover, it was difficult to capture the rich content of images using a small number of key words apart from being an unnatural way of describing images.

Soon it was realized that image retrieval based on the contents is a natural and an effective way of retrieving images. This led to the development of Content-Based Image Retrieval (CBIR). The Content-Based Image Retrieval (CBIR) [8] is aimed at efficient retrieval of relevant images from large image databases based on automatically derived image features. The original Query by Image Content (QBIC) system [3] allowed the user to select the relative importance of color, texture, and shape. The virage system [1] allows queries to be built by combining color, composition (color layout), texture, and structure (object boundary information).

Moments [6] fourier descriptors [5], [11] and chain codes [4], [12] have also been used as features. Jain

and Vailya [7] introduced edge direction histogram (EDH) using the Canny edge operator [2].

The edge orientation autocorrelogram (EOAC) classifies edges based on their orientations and correlation between neighboring edges in a window around the kernel edge [9].

2. SYSTEM ARCHITECTURE

PATSEEK consists of two subsystems- one for Creation of Feature Vector and Image Database and another one for Retrieval of Images similar to the query Image. The components of both subsystems and their interaction are shown in figure 1 (next page).

In order to add the feature vectors and images to the database, the user interacts with the system through its graphical user interface and provides a set of keywords. A snapshot of the user interface for specifying the search criteria for the patents to be grabbed is shown in figure 2 (next page). The patents are grabbed from the USPTO website. The image grabber searches the patent database and grabs the image pages from the patents that satisfy the search criteria.

A page image may contain more than one individual image. To extract the individual images from these page images, we need to identify the connected components or blocks. In these pages, the connectivity is present at a very gross level and an individual image is well separated in both directions from other images. The occasional captions are insignificant compared to the images and can be treated as noise. Our experiment shows that these captions do not change the feature vector of an image in any significant way. To identify the connected components, we start scanning a page image from the first pixel row. If a row has no black pixel, the row is discarded from further consideration. If we find a row that contains at least one black pixel, the position of the row is recorded as the potential start

of a connected component. We continue the horizontal scan as long as we continue to find rows that contain at

least one black pixel.

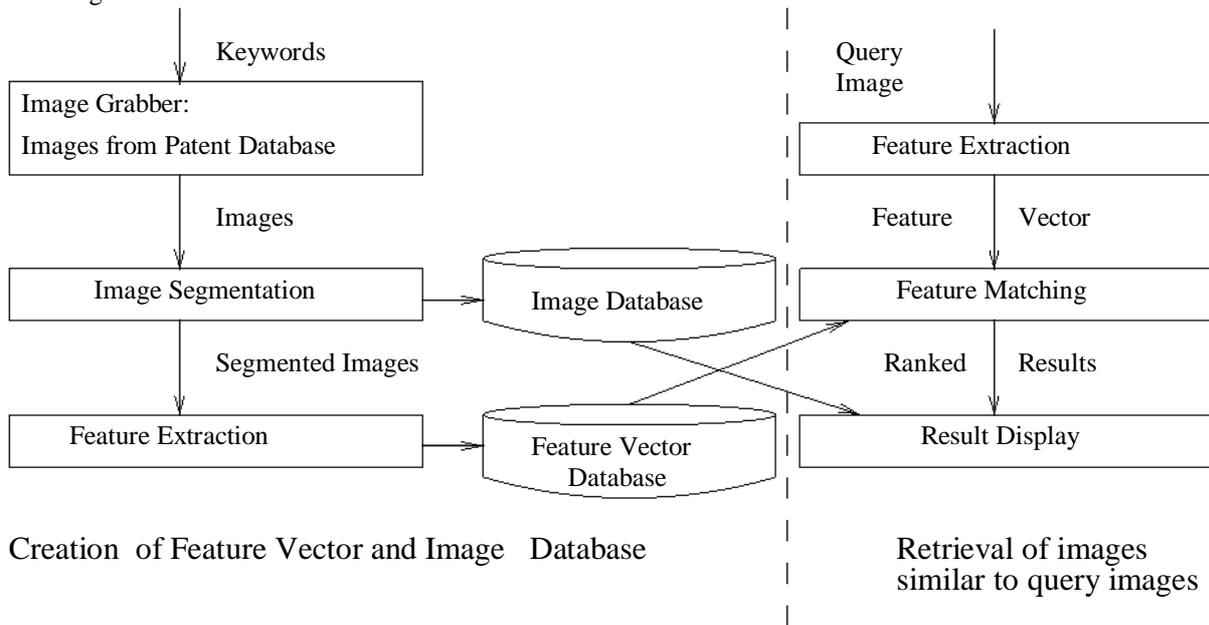


Figure 1. System Architecture of PATSEEK

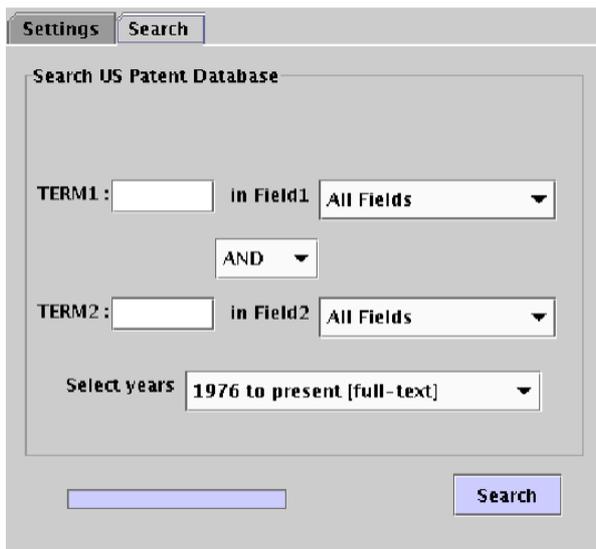


Figure 2. Patent Grabber

If we find enough (design parameter) contiguous horizontal rows that contain no black pixels, we have reached the end of the previous block if any. We continue horizontal scan till the end of the page image. For each horizontal block identified, we scan it vertically to segment it vertically using vertical threshold. These thresholds are set to 1 mm. A block that is less than 5 square cm is discarded as noise.

A document containing three images and their corresponding blocks that have been identified are shown in figure 3.

The separated images are stored in the image database along with patent number and the page number within the patent where these images were found.

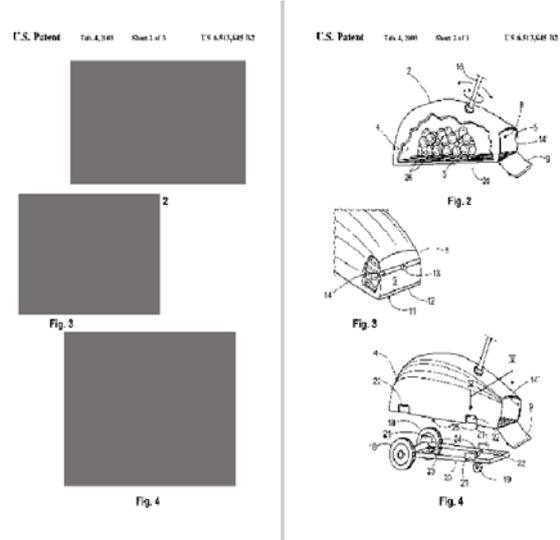


Figure 3. Blocks Identified for Individual Images

The graphic content are then used to calculate the image feature vector. We selected EOAC for our work because it is computationally inexpensive and independent of translation and scaling. Also, the size of the feature vector is small, 144 real numbers. We computed magnitude and gradient of the edges using Canny edge operator.

The edges that have less than 10 percent of the maximum possible intensity are ignored from further consideration.

The gradient of edges is then used to quantize edges into 36 bins of 5 degrees each. The edge orientation autocorrelogram is then formed which is a matrix, consisting of 36 rows and 4 columns. Each element of

this matrix indicates the number of edges with similar orientation. Columns 1, 2, 3 and 4 give the number of

edges that are 1, 3, 5 and 7 pixels apart. Each row corresponds to 5 degrees bin.

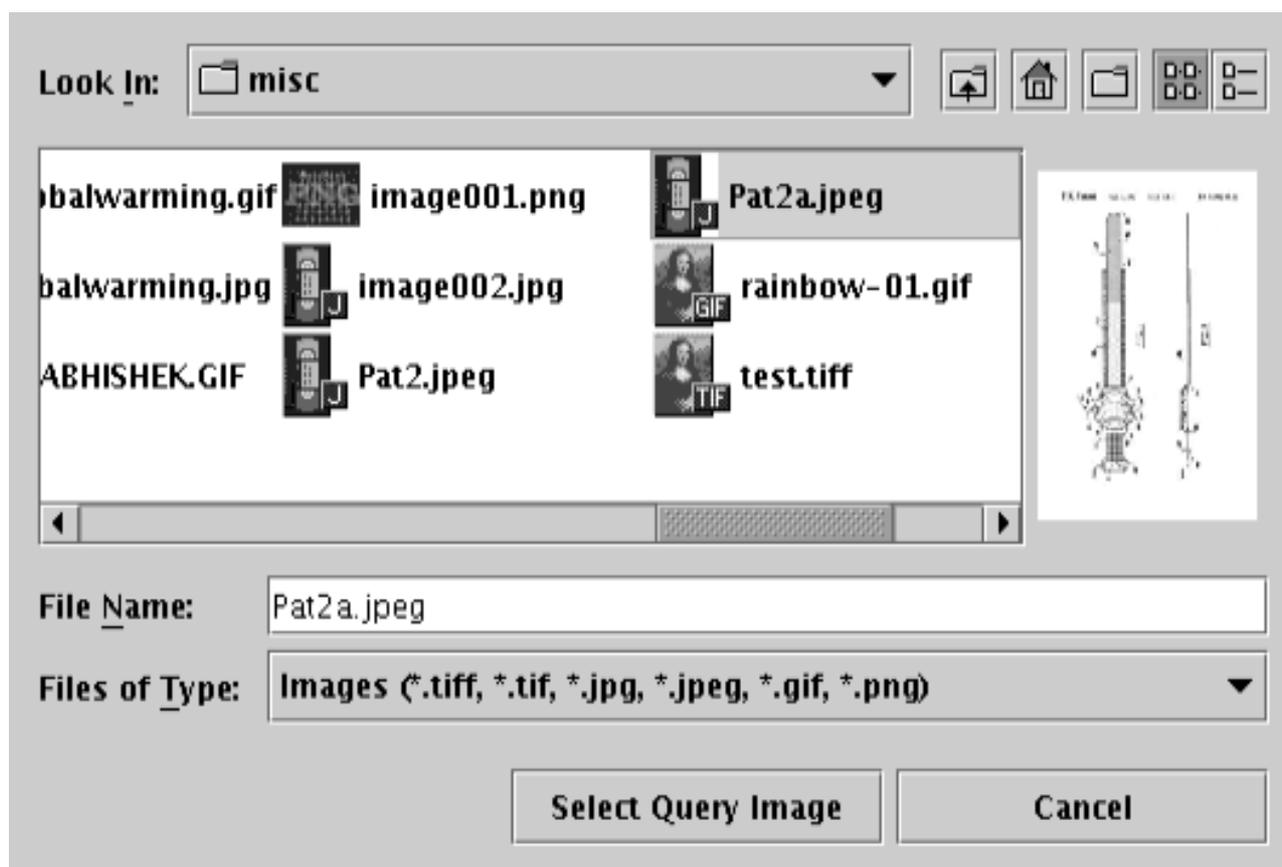


Figure 4. Query Image Selection

Two edges with k pixel distance apart are said to be similar if the absolute values of their orientations and amplitude differences are less than an angle and an amplitude threshold value, respectively [5]. These are user defined thresholds. The edge orientation autocorrelogram is stored in the database along with patent number and the page number within the patent where this image was found.

Feature vectors for the images are stored in an RDMBS table. We have created the feature vector database on Oracle as well as on Mysql.

3. THE DATABASE RETRIEVAL SYSTEM

For querying the database for images similar to a query image, the user interacts with the system through a graphical user interface (shown in figure 4). The user can specify the query image by providing its name and path. The image can be in any of the popular format such as tiff, gif, jpeg etc.

PATSEEK gives user an opportunity to specify the rotation angle for the query image. The angle may range between 0 degrees and 180 degrees.

We have used L1 and L2 distance measures to select 12-nearest neighbors and both have given almost identical results. The top twelve images, ranked on the basis of the distance are displayed as thumbnails along with the respective distance.

The graphical user interface displays the query image and the results for browsing to the user. A snapshot of the user interface is shown in the figure 5.

4. EXPERIMENTS

In this section, we report experimental results. All experiments were performed on an Intel Pentium IV Processor 2.4 GHz with 512 MBytes of RAM. The system was implemented in Java (Sun JDK 1.4.1). The feature vector database was initially created on Oracle and later on moved to MySQL Ver 12.21. We did not use the client of Oracle. We implemented the front-end using Java and its utilities.

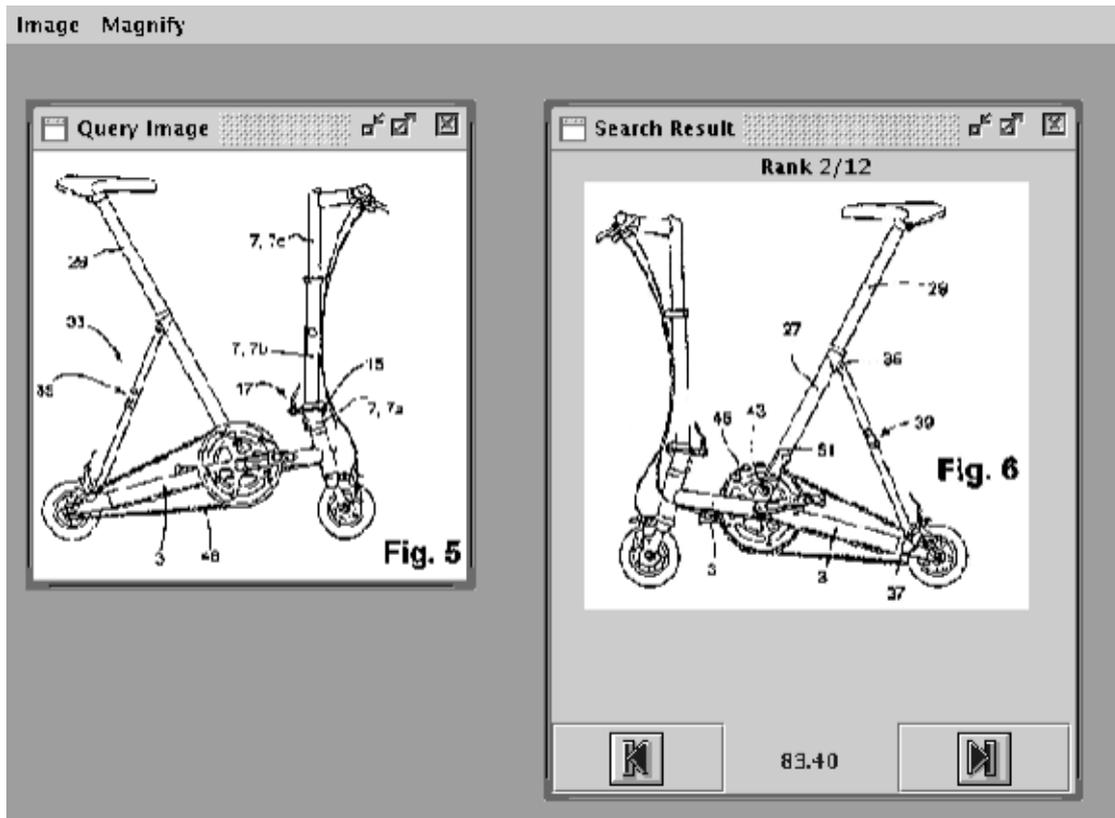


Figure 5. The User Interface for query result navigation

Our database contains approximately 200 images that have been picked up from the patent database of United States Patent Office. For the performance evaluation, we have arbitrarily chosen 15 images from our collection. For each query image, a set of relevant images in the database have been manually identified. An ideal image retrieval system is expected to retrieve all the relevant images. One of the popular measures is *recall rate* [10] that is the ratio of number of relevant images retrieved and total number of relevant images in the collection. The *precision rate* is the ratio of the number of relevant images retrieved and total number of images in the collection.

Minkowski-form distance is used assuming that each dimension of image feature vector is independent of each other and is of equal importance. L1 and L2 (also called Euclidean distance) are some of the widely used Minkowski-form distance measures.

We have compared the performance of two similarity measures, L1 and L2 distance, in our experiments. For our experiment, we calculated precision and recall rate for each image. A recall rate of 100% for 61% of query images and an average 32% recall rate for rest of the images were observed. The precision rate varied between 10% and 35%. Precision rate greatly depend on the number of the images in the database. Graph 1 and graph 2 show the precision rates and recall rates for 15 query images. The image shown in figure 6 is one of the images selected for the query image shown on the left in figure 4 but it is not one of the top

ranking images. For the same query image, when we specify the angle of rotation as 180 degrees, the ranking of this image improves considerably.

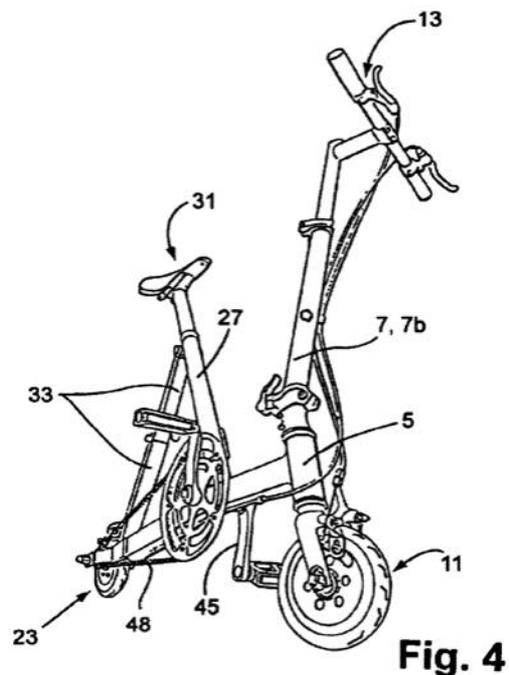
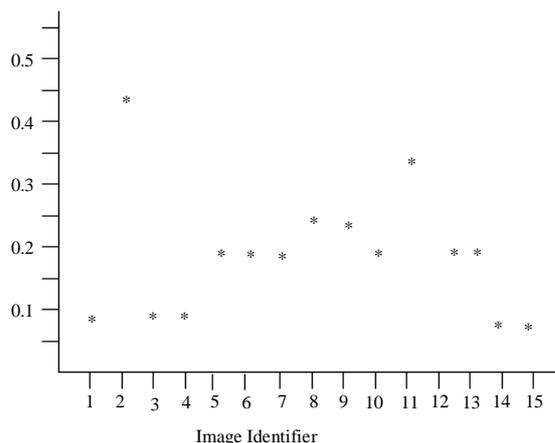
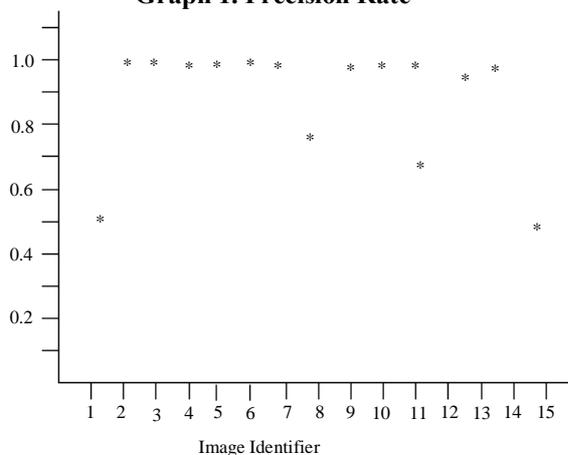


Figure 6. An image obtained by rotating the query image



Graph 1. Precision Rate



Graph 2. Recall rate

5. CONCLUSION

In this paper, we have described PATSEEK that is an image retrieval system for the US patent database. The system has shown good performance and can be effectively utilized to locate similar patents before issuing a new patent by the patent office. A researcher or developer can locate all the images similar to the image in his document before filing for a patent. PATSEEK can compliment a text based search system.

The image feature vector database for PATSEEK is expected to grow and the retrieval system must give real-time performance even when database is large. At present, we have not made any effort to optimize the speed. The total time elapsed from the moment a query

image was given and to the moment relevant images are retrieved was about 90 seconds. We plan to use multidimensional indexing to cut down the retrieval time.

REFERENCES

- [1] Bach, J. R., et al., "The virage image search engine: an open framework for image management", *Proc. SPIE: Storage and retrieval for Still Image and Video Databases IV*, vol 2670, pp 76-87, 1996.
- [2] Canny, J., "A computational approach to edge detection", *IEEE trans. Pattern Analysis and Machine Intelligence*, PAMI-8, pp 679-698, 1986.
- [3] Flickener, M., et al., "Query by image and video content: the QBIC system", *IEEE Computer*, vol 28, pp 23-32, 1995.
- [4] Freeman, H., "On the encoding of arbitrary geometric configurations", *IRE Trans. on Electronic Computers*, vol EC-10, pp 260-268, 1961.
- [5] Gonzalez R.C. and Wints P., *Digital Image Processing*, Addison-Wesley Reading, MA, 1992.
- [6] Hu, M. K., "Visual pattern recognition by moments invariants", *IRE Transactions on Information Theory*, IT-8, pp. 179-187, 1962.
- [7] Jain, A.K. and Vailaya, A., "Image Retrieval using Color and Shape", *Pattern Recognition*, vol 29, pp 1233-1244, August 1996.
- [8] Kato, T., "Database architecture for content-based image retrieval in Image Storage and Retrieval Systems" (Jambardino A and Niblack W eds), *Proc SPIE 2185*, pp 112-123, 1992.
- [9] Mahmoudi, F., Shanbehzadeh J., Eftekhari A.M., and Soltanian-Zadeh H., "Image retrieval based on shape similarity by edge orientation autocorrelogram", *Pattern Recognition*, vol 36, pp 1725-1736, 2003.
- [10] Muller, H., Mullerm, W., McGSquire, D., Marchand-Maillet, S., and Pun, T., "Performance evaluation in content-based image retrieval: overview and proposals", *Pattern Recognition Letters*, vol 22, pp 593-601, 2001.
- [11] Persoon, E. and Fu, K. S., "Shape discrimination using Fourier descriptors", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 8, pp 388-397, 1986.
- [12] Zhang, D. and Lu, G., "Review of shape representation and description techniques", *Pattern Recognition*, 37(1), pp 1-19, 2004.

Privacy-Preserving Collaborative Association Rule Mining

Justin Zhan¹, Stan Matwin¹, Nathalie Japkowicz¹, LiWu Chang²

¹ School of Information Technology & Engineering, University of Ottawa, Canada

² Center For High Assurance Computer Systems, Naval Research Laboratory, USA
{zhizhan, stan, nat}@site.uottawa.ca, lchang@itd.nrl.navy.mil

ABSTRACT

In recent times, the development of privacy technologies has promoted the speed of research on privacy-preserving collaborative data mining. People borrowed the ideas of secure multi-party computation and developed secure multi-party protocols to deal with privacy-preserving collaborative data mining problems. Random perturbation was also identified to be an efficient estimation technique to solve the problems. Both secure multi-party protocol and random perturbation technique have their advantages and shortcomings. In this paper, we develop a new approach that combines existing techniques in such a way that the new approach gains the advantages from both of them.

Keywords: Privacy, data mining, association rule mining, secure multi-party computation.

1. INTRODUCTION

Business successes are no longer the result of an individual toiling in isolation; rather successes are dependent upon collaboration, team efforts, and partnership. In the modern business world, collaboration becomes especially important because of the mutual benefits it brings. Specially, the explosion in the availability of various kinds of data has triggered tremendous opportunities for collaboration. This paper studies a very specific collaboration that becomes more and more prevalent. The problem is the collaborative data mining. The objective of our research is to develop technologies to enable multiple parties to conduct data mining collaboratively without disclosing their private data. Since its introduction [7], privacy-preserving collaborative data mining problem has attracted many researchers. The approaches applied to solve the problem can be categorized into two aspects: one is based on secure multi-party computation [3, 10]; the other is based on randomization techniques [1, 4,5]. The first approach can achieve precise results but its computation cost is usually much greater than that of the second approach. The second approach is computationally efficient but its accuracy is not as good as that of the first approach. This motivates us to find a combination in which the two approaches are integrated. The combined approach gains advantage of good accuracy that the first approach has, without losing benefit of high efficiency that the second approach has. In this paper, a secure integrated protocol is presented to deal with collaborative data mining problems. Collaborative data mining includes a number of different tasks, such as collaborative association rule mining, classification, and clustering, etc. This paper studies collaborative association rule mining problem. The goal is to discover meaningful association rules among the attributes of a large quantity of data contributed by all the collaborative parties.

The paper is organized as follows: Section 2 discusses the related work. Section 3 formally defines the

privacy-preserving multi-party association rule mining problem. We develop a secure protocol in Section 4. We describe our experimental results in Section 5. Further discussion is provided in Section 6. Section 7 concludes the paper and lays out the future work.

2. RELATED WORK

2.1 Secure Multi-party Computation

Briefly, a Secure Multi-party Computation (SMC) problem deals with computing any function on any input, in a distributed network where each participant holds one of the inputs, while ensuring that no more information is revealed to a participant in the computation than can be inferred from that participant's input and output [6]. It has been proved that for any polynomial function, there is a secure multi-party computation solution [6]. The approach used is as follows: the function F to be computed is first represented as a combinatorial circuit, and then the parties run a short protocol for every gate in the circuit. Every participant gets corresponding shares of the input wires and the output wires for every gate. This approach, though appealing in its generality and simplicity, is highly impractical for large datasets.

2.2 Privacy-Preserving Data Mining

In early work on such a privacy-preserving multi-party data mining problem, Lindell and Pinkas [7] propose a solution to privacy preserving classification problem using oblivious transfer protocol, a powerful tool developed by secure multi-party computation (SMC) research. Other techniques based on SMC were then proposed in [3,10]. Randomization approaches was firstly proposed to solve privacy-preserving data mining problem by Agrawal and Srikant in [1]. People proposed more random perturbation based techniques to tackle the problem in [1,4,5]. The techniques based on SMC have characteristics that mining results will be the

same as original results without considering data privacy. In contrast, mining results using techniques based on randomization will be the estimation of original results but performance is usually better than SMC-based techniques. In this paper, we propose an approach which requires integrated use of two existing approaches. Through integration of two approaches, we can gain advantages of both approaches.

3. INTRODUCING THE PROBLEM

We consider the scenario where multiple parties, each having a private data set denoted by D_1, D_2, \dots , and D_n respectively where n is the total number of parties, want to collaboratively conduct association rule mining on the union of their data sets. Because they are concerned about their data privacy, neither party is willing to disclose its raw data set to others. Specially, we consider the scenario where data sets of collaborative parties are binary and vertically partitioned [10]. In other words, the identities of the i th (for $i \in [1, N]$) record in D_1, D_2, \dots , and D_n are the same, and D_1, D_2, \dots , and D_n contain the same number of records (we use N to denote the total number of records for each data set).

4. SECURE MULTI-PARTY ASSOCIATION RULE MINING PROTOCOL

4.1 The Commodity Server Model

The commodity server model was proposed in [2]. The collaborative parties could send requests to the commodity server and receive data from the server, but the commodities must be independent of the parties' private data. The purpose of the commodities is to help the parties to conduct their desired computations. The commodity server is semi-trusted in the following senses: (1) It should not be possible to derive private information of data from the parties; it should not learn computation result either. (2) It should not collude with all the parties. (3) It follows the protocol correctly. Because of these characteristics, we say that it is a semi-trusted party. In real world, finding such a semi-trusted party is much easier than finding a trusted party. Based on this commodity server model, we study privacy-preserving multi-party association rule mining problem. In our solution, the commodities are randomly generated permutation functions [8]. Our goal is to permute all the parties' data in such a way that a designated party (data collector) can conduct association rule mining on permuted data and obtain the same mining results. The challenge is to prevent the designated party from knowing how the other parties' data are permuted.

4.2 Notations

For convenience, we define the following intuitive

notations:

1. Let V represent a permutation function that is used to permute the order of the columns in a data set. We call it the column permutation function. We use V^{-1} to represent the inverse of V .
2. Let H represent a permutation function that is used to permute the order of the rows in a data set. We call it the row permutation function. We use H^{-1} to represent the inverse of H .
3. Let $VH = (V, H)$ represent a permutation function that contains a row permutation H and a column permutation V . VH means that we firstly apply H on the data set, V is then applied. We use $(VH)^{-1}$ to represent the inverse of VH .
4. We use \Rightarrow as the permutation sign. For instance, $R_1 \Rightarrow R_3$ means that Row 1 will be permuted to Row 3 and $C_1 \Rightarrow C_4$ means that Column 1 will be permuted to Column 4.

4.3 Privacy-Preserving Multi-Party Association Rule Mining Protocol

At the beginning of our protocol, a data collector is randomly selected from all the collaborative parties. For simplicity, let's assume Party n is selected as data collector. The commodity server then generates a set of permutation functions satisfying some properties and sends them to the collaborative parties. Party 1, 2, ..., $n-1$ permute their data by themselves and send permuted data to Party n . Since Party n also contributes its data set to the mining computation, its data must also be permuted similarly, i.e., the row permutation effects have to be the same. To avoid information leak, the permutation of Party n 's data has to be handled differently. In our protocol, the permutation of Party n 's data is done by itself together with another party (Party i) who is randomly selected from Party 1, 2, ..., $n-1$. Party i and Party n will together permute the Party n 's data such that the permuted data have the same row permutation as other permuted data sets. To prevent Party n from conducting possible attacks, multi-variant randomized response technique [4] is applied on the data of Party n .

Multi-Party Mining Protocol

Step I: Permutation Function Generation

1. The Commodity Server (CS) generates a set of column permutation functions $V_1, V_2, V_3, \dots, V_{n-1}$ and a row permutation function H . It then sends V_1H to Party 1, V_2H to Party 2, ..., and $V_{n-1}H$ to Party $n-1$.
2. The CS generates the other two sets of permutation functions $V_n'H$ and $V_n''H$. It then sends $V_n'H$ to Party n , $V_n''H$ to Party i where Party i is randomly selected from Party 1, Party 2 ... and Party $n-1$.

3. The permutation functions have the following property: $V_1H, V_2H, \dots, V_{n-1}H$ and $V_n'H''V_n'H'$ have the same row permutation effects.

Definition 1 We say two permutation functions PF_1 and PF_2 have the same row permutation effects if $PF_1(D_1)$ and $PF_2(D_2)$ have the same row permutations.

Step II: Data Permutation

Sub-step 1: Permute $D_1, D_2, \dots,$ and D_{n-1}

1. Party 1 applies V_1H on its data set D_1 to get a permuted data set PD_1 , then sends PD_1 to Party n.
2. Party 2 applies V_2H on D_2 to obtain PD_2 , and then sends it to Party n.
3. Repeat the above process until Party n-1 applies $V_{n-1}H$ on its data set D_{n-1} and obtain PD_{n-1} , then sends it to Party n.

Sub-step 2: Permute D_n

1. Party n applies $V_n'H'$ on D_n to obtain PD'_n , then sends it to Party i.
2. Party i applies $V_n''H''$ on PD'_n to obtain PD''_n , then randomizes PD''_n using the multi-variant randomized response technique described in Section 4.5, finally obtains PD_n and sends it to Party n (We will explain why Party i randomizes PD''_n).

Sub-step 3: Combine the Permuted Data Sets

1. Party n concatenates all the data sets $PD_1, PD_2, \dots, PD_{n-1}$ and PD_n such that the *i*th row in the PD_1, PD_2, \dots and PD_n becomes the *i*th row in the joint data set PD.

Step III: Mining

1. Party n conducts the association rule mining on the joint data sets PD and gets the permuted association rules denoted by PAR_n . Since the data collector's data are randomized, the standard association rule mining algorithm cannot be employed. Section 4.7 shows how to conduct association rule mining on data sets which contain both non-randomized data (e.g., the data of Party 1, 2... n-1) and randomized data (e.g., the data of Party n).

Step IV: Final Association Rules

Sub-step 1: Find the Final Rules Containing the Attributes from Party 1, Party 2, ..., and Party n-1

1. Party n sends the attributes, which the permuted association rules contain, to the corresponding parties (from Party 1 to Party n-1) who send the name of the

original attributes to Party n. For example, if the permuted rules contains a rule $PD_3(A_1) \Rightarrow PD_5(A_4)$ with $PD_3(A_1)$ from Party 3 and $PD_5(A_4)$ from Party 5, Party n then sends $PD_3(A_1)$ to Party 3 and $PD_5(A_4)$ to Party 5. Without loss of generality, assume the original attributes of $PD_3(A_1)$ and $PD_5(A_4)$ are $D_3(A_8)$ and $D_5(A_3)$ respectively, they then send the attribute names of $D_3(A_8)$ and $D_5(A_3)$ to Party n. Note that the attribute indices are used in the protocol, e.g., $PD_3(A_1)$ denotes the first attribute in the permuted data set of Party 3. However, the data collector does not know whether $PD_3(A_1)$ stands for *bread*, *butter*, or others. After the parties remove the column permutation functions and obtain the original attribute indices, e.g., Party 3 gets $D_3(A_8)$, they will not send the original attribute indices (e.g., $D_3(A_8)$) to the data collector and only sends the real attribute name that the original attribute indices denote to Party n. For instance, if the $D_3(A_8)$ stands for *bread*, it sends *bread* to the data collector. The purpose of the above process is to prevent the data collector from inferring other parties' column permutation functions based on the resultant rules.

Sub-step 2: Find the Final Rules Containing Party n's Attributes

1. Party n sends its permuted attributes that the permuted association rules contain, to Party i who applies the inverse of V''_n on the permuted attributes and obtains some attributes. We call them the middle permuted attributes since they are still in permuted form (e.g., permuted by V'_n).
2. Party i sends the middle permuted attributes to Party n who applies the inverse of V'_n on them and obtains the original Party n's attributes.

Sub-step 3: Find the Final Rules

After obtaining all the original attributes of the permuted attributes that the permuted association rules contain, Party n gets the final association rules and shares them with all the other parties.

4.4 Why Party i Randomizes the Party n's Data

Assume that after Party i applies $V''_n H''_n$ on the data $V'_n H'_n (D_n)$ obtained from Party n, Party i directly sends $V''_n H''_n V'_n H'_n (D_n)$, without randomization, back to Party n. Since the total number of records N is usually much larger than the total number of attributes M in a large-scale data set, the probability of the sum of each column being unique is high. Thus, Party n is likely to find out the column permutation function V''_n by comparing $V'_n H'_n (D_n)$ with $V''_n H''_n V'_n H'_n (D_n)$ using the sum of each column. After finding V''_n , she can

remove V''_n from $V''_n H''_n V''_n H''_n (D_n)$. If Party n can somehow find out H''_n , Party n then knows $V''_n H''_n$. In other words, she knows $V''_n H''_n V''_n H''_n (D_n)$. Since all the parties' data are permuted in a way with the same row permutation effects and Party n also obtains other parties' permuted data, Party n can find out other parties' row permutation function H . H''_n is likely to be disclosed for any row of D_n that has a unique pattern. To avoid this problem, we let Party i randomize $V''_n H''_n V''_n H''_n (D_n)$ before sending it to Party n. Even if the Party n's data has the above special characteristics, Party n cannot exactly identify the Party i's permutation function $V''_n H''_n$ because of randomization. The challenge is how Party n conducts association rule mining after obtaining data set PD . Obviously, Party n cannot apply the traditional association rule mining algorithm [9], since part of the data set are not randomized and part of them are randomized. We will show how Party n conducts the association rule mining on the hybrid data set PD in Section 4.7.

4.5 How Party i Randomizes The Party n's Data

We propose to use randomization technique [4] to disguise $V''_n H''_n V''_n H''_n (D_n)$. The basic idea is that, for each record, we keep the same values with a predefined probability q ; we reverse the values with the probability of $1-q$. For example, assume there are 3 attributes for a record and the values for the record is $(A_1=1, A_2=1, A_3=0)$. We generate a random number between 0 and 1. If the number is less than q , the record is kept the same, that is $(A_1=1, A_2=1, A_3=0)$. If the number is greater than q , the record becomes $(A_1=0, A_2=0, A_3=1)$. For each record, we generate a random number to decide whether the record keeps the same values or being reversed. Finally, Party i obtains the data set PD_n and sends it to Party n. To conduct association rule mining on the randomized data set, we need to compute $P(A_1=1, A_2=1, A_3=0)$. To simplify our presentation, we use $P(110)$ to represent $P(A_1=1, A_2=1, A_3=0)$, $P(001)$ to represent $P(A_1=0, A_2=0, A_3=1)$. Because the contributions to $P^*(110)$ and $P^*(001)$ partially come from $P(110)$, and partially come from $P(001)$, we can derive the following equations:

$$P^*(110) = P(110) \cdot q + P(001) \cdot (1-q)$$

$$P^*(001) = P(001) \cdot q + P(110) \cdot (1-q)$$

By solving the above equations, we can get $P(110)$. The general model is described in the following:

$$P^*(E) = P(E) \cdot q + P(\bar{E}) \cdot (1-q)$$

$$P^*(\bar{E}) = P(\bar{E}) \cdot q + P(E) \cdot (1-q)$$

where E is a logical expression. In the above example, $E=(A_1=1, A_2=1, A_3=0)$ and $\bar{E}=(A_1=0, A_2=0, A_3=1)$. To evaluate the proposed randomization approach, we conduct extensive experiments in Section 5.

4.6 How Can We Obtain the Final Association Rules from The Permuted Association Rules

In the mining step, Party n obtains the permuted association rules PAR_n . To obtain the final association rules from PAR_n , the parties need to apply the inverse of their column permutation functions on the permuted association rules. In the following, we use a simple example to illustrate the computation. Without loss of generality, we assume the original data, permuted data and permutation functions are shown in Fig. 1. Based on the permuted data, suppose that we obtain the permuted association rules as shown in Fig. 2. We can then apply the inverse of the column permutation functions and get the final association rules (Fig. 2).

4.7 Multi-Party Association Rules Mining On Randomized Data

The following is the procedure for mining association rules on $PD_1 \cup PD_2 \cup \dots \cup PD_n$. It differs from traditional association rule mining procedure [9] in the computation of the frequency count (denoted as $c.count$). When the data are not randomized, we can easily compute the $c.count$, but when part of the data are randomized using the multi-variant randomized response techniques, computing it becomes non-trivial.

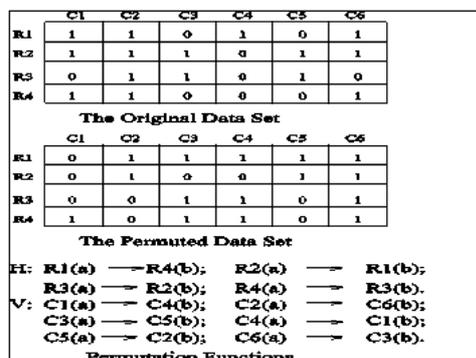


FIGURE 1

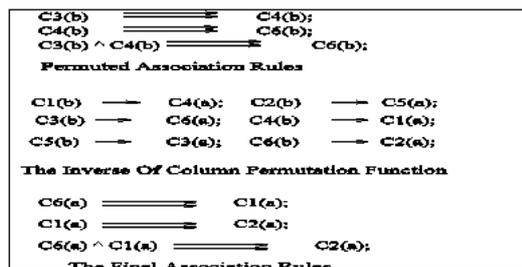


FIGURE 2

For the part which is randomized, we do not know whether a record in the data set is true information or false information. Therefore, we cannot directly compute $c.count$ on the randomized data. In the following, we will show how to compute $c.count$ when some attributes are randomized. For example, we want to compute $c.count$ for A_1 , A_2 , and A_3 . Let's assume A_1 and A_2 are not randomized and A_3 is randomized. We can use the following estimation model.

$$P^*(A_1A_2A_3) = P(A_1A_2A_3) \cdot q + P(A_1A_2\bar{A}_3) \cdot (1-q)$$

$$P^*(A_1A_2\bar{A}_3) = P(A_1A_2\bar{A}_3) \cdot q + P(A_1A_2A_3) \cdot (1-q)$$

In general, when part of the attributes (denoted by RE) are randomized and the other part (denoted by NE) is not randomized, then the estimation model becomes:

$$P^*(RE \wedge NE) = P(RE \wedge NE) \cdot q + P(\bar{RE} \wedge NE) \cdot (1-q)$$

$$P^*(\bar{RE} \wedge NE) = P(\bar{RE} \wedge NE) \cdot q + P(RE \wedge NE) \cdot (1-q)$$

where \wedge is logical and operator. $P^*(RE \wedge NE)$ and $P^*(\bar{RE} \wedge NE)$ can be obtained from the randomized data set and q is known (With the knowledge of q , Party n only knows the probability of each value of PD_n comes from the values of $V_n H_n V_n H_n (D_n)$). We can solve the above equation and get $P(RE \wedge NE)$. We can then compute $c.count = P(RE \wedge NE) * N$.

5. EXPERIMENTAL RESULTS

To evaluate the effectiveness of our proposed scheme, we conduct association rule mining on the original data using the traditional association rule mining algorithm [18] and get a benchmark rule sets T_D . We then induce our rule set T_G from PD using our modified algorithm. We test the difference of these two sets. Since Party n's data are randomized by the multi-variant randomized response technique, we modified the association rule mining algorithm to handle the randomized data based on our proposed methods in Section 4.7. We run this modified algorithm on the randomized data, and find a set of association rules (e.g., our rule set T_G). We also apply the original association rule mining algorithm to the original data set and find the other set of association rules. We then compute the error which is defined as the number of rules that the benchmark rule set T_D contains but our rule set T_G does not contain. The error rate is defined as the error divides the total number of rules in the benchmark rule set T_D . In this section, we will show the experimental results on two real-life data sets.

5.1 Data Setup

We conducted experiments on two real life data sets. We obtain the data sets from the UCI Machine Learning

Repository. The first data set is *Adult*. The second data set is *Breast-Cancer*.

5.2 Experimental Steps

Our experiments consist of the following steps:

Step I: Since we assume that the data set contains only binary data, we firstly transformed the original non-binary data to the binary (D). We split the value of each attribute from the median point of the range of the attribute.

Step II: We use D and the original association rule mining algorithm to build a rule set T_D .

Step III: We randomly select 10% number of attributes from D. For $q = 0, 0.1, 0.2, 0.3, 0.4, 0.51$ (We use 0.51 instead of 0.5. Because when $q = 0.5$, estimation equation doesn't have a unique solution. We call this point an exception point in our approach.), 0.6, 0.7, 0.8, 0.9, 1, we conduct the following 4 steps:

1. Randomization: We create a randomized data set G. For each record in the training data set D, we generate a random number r from 0 to 1 using uniform distribution. If $r \leq q$, we copy the record to G without any change; if $r > q$, we copy the opposite values of selected attributes in the current record to G and we copy the original values of non-selected attributes in the current record to G. For instance, suppose there are three attributes (e.g., A_1 , A_2 , and A_3) in D and the original values for them are $A_1=0$, $A_2=0$ and $A_3=0$. If the selected attributes are A_1 and A_2 . When $r > q$, the values for these two attributes will be $A_1=1$ and $A_2=1$ in the current record of G, but the value for A_3 is still 1. We perform this randomization step for all the records in the training data set D and generate the new data set G.
2. Build a rule set: We use the data set G and our modified association rule mining algorithm to build our rule set T_G .
3. Compute the error: We compare T_D and T_G , and compute the error rate.
4. Repeating: We repeat the above steps for 50 times, and get E_1, E_2, \dots, E_{50} . We then compute the mean of these 50 error rate scores.

Step IV: We randomly select 20% and 50% number of attributes respectively from D. We repeat Step III (1-4). Since there are at least two parties in the multi-party data mining, Party n has at most 50% number of attributes in the combined data set PD if we assume that each party's data have the same number of attributes. Therefore, we didn't conduct experiments when more than 50% attributes are randomized.

θ	0	0.1	0.2	0.3	0.4	0.51	0.6	0.7	0.8	0.9	1
Error Rate	0	0	0	0	0.023	0.256	0.024	0	0	0	0

(a) 10% Attributes Are Randomized

θ	0	0.1	0.2	0.3	0.4	0.51	0.6	0.7	0.8	0.9	1
Error Rate	0	0.12	0.12	0.2	0.352	0.336	0.24	0.24	0.12	0.06	0

(b) 20% Attributes Are Randomized

θ	0	0.1	0.2	0.3	0.4	0.51	0.6	0.7	0.8	0.9	1
Error Rate	0	0.16	0.16	0.32	0.326	0.376	0.247	0.24	0.168	0.12	0

(c) 50% Attributes Are Randomized

FIGURE 3

5.3 Result Analysis

Fig. 3 shows the mean of error for *Adult* data set with the minsup=0.6. Fig. 4 depicts the mean of error for *Breast-Cancer* data sets with the minsup=0.45. We can see when $q=1$ and $q=0$, the results are exactly the same as the results when the original association rule mining algorithm is applied. This is because when $q=1$, the randomized data sets are exactly the same as the original data set D; when $q=0$, the randomized data sets are exactly the opposite of the original data set D. In both cases, our algorithm produces the accurate results comparing to the original algorithm. However, the privacy is not preserved in either case because an adversary can know the real values of all the records provided that she knows the q value. When q moves from 1 and 0 towards 0.5, the mean of error has the trend of increasing, and the probability of disclosing the original data is decreasing. Therefore the privacy level of the data increases. When more and more data are randomized, the mean of error will increase.

θ	0	0.1	0.2	0.3	0.4	0.51	0.6	0.7	0.8	0.9	1
Error Rate	0	0	0	0	0.04	0.044	0.03	0	0	0	0

(a) 10% Attributes Are Randomized

θ	0	0.1	0.2	0.3	0.4	0.51	0.6	0.7	0.8	0.9	1
Error Rate	0	0.043	0.043	0.044	0.047	0.065	0.042	0.041	0.039	0.024	0

(b) 20% Attributes Are Randomized

θ	0	0.1	0.2	0.3	0.4	0.51	0.6	0.7	0.8	0.9	1
Error Rate	0	0.088	0.107	0.113	0.116	0.183	0.169	0.104	0.077	0.05	0

(c) 50% Attributes Are Randomized

FIGURE 4

6. DISCUSSION

In this paper, we combine the secure multi-party computation approach with the random perturbation approach to tackle the problem of privacy-preserving collaborative association rule mining. As shown in previous sections, all parties' data are permuted, but Party n's data are also randomized. As a result, the

mined rules containing Party n's attributes are degraded with a certain level of inaccuracy. On the other hand, the rules that do not contain Party n attributes will be 100% accurate comparing to the results obtained by applying the traditional association rule mining algorithm [9] on the non-randomized and non-permuted data. Therefore, the results will be the tradeoff between the SMC-based approach and the randomization approach. Our secure multi-party mining protocol gains the advantage of good accuracy that the SMC-based approach has, without losing the benefit of high efficiency that the randomization approach has. Although we deal with the association rule mining problem in this paper, the proposed approach can be extended to solve other data mining problems, e.g., decision tree, Bayesian network classification, etc. To improve privacy level, we can also use multi-group randomized response [11] to randomize the Party n's data. We need also point out that, Party n can apply maximum likelihood estimation method to attack H^n when the distribution of rows of D_n is highly skewed, and the number of records of D_n is small. Therefore, the data distribution should be considered when selecting the data collector.

7. CONCLUSION

We have presented a secure protocol to let multiple parties to find association rules on the joint data sets without comprising their data privacy. In the protocol, the data collector is firstly selected. All the parties except for the data collector permute their data using the permutation functions obtained from the commodity server. To prevent the data collector from conducting possible attacks on other parties' permutation functions, the multi-variant randomized response technique is utilized to randomize data collector's data. Finally, we conduct a set of experiments to evaluate the effectiveness of our proposed scheme.

In this paper, we only consider the semi-trust case in which all the collaborative parties honestly follow the protocol. The more challenging scenario is the malicious case where the collaborative parties may cheat with each other. As future work, we will develop the protocol to deal with the malicious case.

REFERENCES

[1] R. Agrawal and R. Srikant, Privacy-preserving data mining, In proceedings of the ACM SIGMOD conference on management of data, Dallas, Texas, USA, 2000.
 [2] D. Beaver, Commodity-based cryptography (extended abstract), In proceedings of the twenty-ninth annual ACM symposium on theory of computing, EL Paso, TX USA, May 4-6, 1997.
 [3] W. Du and Z. Zhan, Building decision tree classifier on private data, In workshop on privacy, security and data mining at the 2002 IEEE international

- conference on data mining, Maebashi City, Japan, December 9, 2002.
- [4] W. Du and Z. Zhan, Using randomized response techniques for privacy-preserving data mining, In proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining, Washington, DC, USA, August 24-27, 2003.
- [5] A. Evfimievski, J. Gehrke, and R. Srikant, Limiting privacy breaches in privacy-preserving data mining, In the proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems, San Diego, California, 2003.
- [6] O. Goldreich, Secure multi-party computation (working draft), http://www.wisdom.weizmann.ac.il/home/oded/public_html/foc.html, 1998.
- [7] Y. Lindell and B. Pinkas, Privacy-preserving data mining, Advantages in cryptology-CRYPTO'00, 1880 of lecture notes in computer science, Springer-Verlag: 36-54, 2000.
- [8] M. Luby, Pseudorandomness and cryptographic applications, Princeton University Press, Jan., 1996.
- [9] R. Agrawal, T. Imielinshi and A. Wigderson, mining association rules between sets of items in large databases, In proceedings of ACM SIGMOD conference on management of data, pages:207-216, May, 1993.
- [10] J. Vaidya and C. Clifton, Privacy-preserving association rule mining in vertically partitioned data, In proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining, July 23-26, 2002.
- [11] Z. Zhan, L. Chang and S. Matwin, Privacy-preserving multi-party decision tree induction, In the 18th annual IFIP WG 11.3 working conference on data and application security, Sitges, Catalonia, Spain, 2004.

Privacy-Preserving Data Mining In Electronic Surveys

Justin Zhan, Stan Matwin

School of Information Technology & Engineering, University of Ottawa, Canada
{zhizhan, stan}@site.uottawa.ca

ABSTRACT

Electronic surveys are an important resource in data mining. However, how to protect respondents' data privacy during the survey is a challenge to the security and privacy community. In this paper, we develop a scheme to solve the problem of privacy-preserving data mining in electronic surveys. We propose a randomized response technique to collect the data from the respondents. We then demonstrate how to perform data mining computations on randomized data. Specifically, we apply our scheme to build a Naive Bayesian classifier from randomized data. Our experimental results indicate that accuracy of classification in our scheme, when private data is protected by randomization, is close to the accuracy of a classifier build from the same data with the total disclosure of private information. Finally, we develop a measure to quantify privacy achieved by our proposed scheme.

Keywords: Privacy, Data mining, Randomization

1. INTRODUCTION

Data mining has emerged as a means for identifying patterns and trends from large amounts of data. To conduct data mining computations, we need to collect data first. However, because of privacy concerns, people might decide to selectively divulge information, or give false information, or simply refuse to disclose any information at all. There is research evidence [4] that providing privacy protection measures is a key to the success of data collection.

There are many ways to collect data. For instance, data may be collected using transaction records. This can often be done without people's knowledge, and individuals have no control over what information can be collected. The evolving legal developments will hopefully soon preclude this questionable practice. Another way to collect data is to solicit respondents' responses via surveys, for example, respondents might be asked to rate certain products, or they might be asked whether they have a certain medical condition, etc. The collected data is entered into a database. Although answering survey questions gives respondents control over whether they want to disclose their information or not, privacy concerns might hinder the respondents from telling the truth or responding at all (we will refer to this problem as *respondent privacy in electronic surveys*). How can we improve the chance to collect more truthful data that are useful for data mining while preserving respondents' privacy? How can respondents contribute their personal information without compromising their privacy?

We propose to use *Randomized Response* techniques [9] to solve the problem of respondent privacy in electronic surveys. The basic idea of randomized response is to scramble the data in such a way that data collector cannot tell with probabilities better than a pre-defined threshold whether the data from a respondent contain truthful information about the sensitive, private

information. Although information from each individual respondent is scrambled, if the number of respondents is significantly large, the aggregate information of these respondents can be estimated with reasonable accuracy. Such property is useful for naive Bayesian classification since it is based on aggregate values of a data set, rather than individual data items. The contributions of this paper are as follows: (1) We have modified naive Bayesian classification algorithm [7] to make it work with data disguised by randomized response techniques, and implemented the modified algorithm. (2) We then conducted a series of experiments to measure accuracy of our modified naive Bayesian algorithm on randomized data. Our results show that if we choose appropriate randomization parameters, the accuracy we have achieved is very close to the accuracy achieved by standard, unmodified naive Bayesian classifier on undisguised data. (3) We have developed a method to measure privacy achieved by our proposed approach.

There are some related works [1, 2, 3, 5, 6, 8, 10] in privacy-preserving data mining. We do not discuss them in details because of space.

2. BUILDING NAÏVE BAYESIAN CLASSIFIERS USING MULTI-VARIANT RANDOMIZED RESPONSE TECHNIQUES

Randomized Response techniques were first introduced by Warner [9] to solve the following survey problem: to estimate the percentage of respondents in a population that has attribute A, queries are sent to a group of respondents. Since the attribute A is related to some confidential aspects of human life, respondents may decide not to reply at all or to reply with incorrect answers. For the purpose of this discussion, we will distinguish two types of questions in a survey: questions about the respondent's private information, and questions about the respondent's personal information. Both kinds of information refer to the attributes of the

respondent. The private information is an attribute the respondent would rather not disclose, including its probability distribution (e.g., whether the respondent has a certain medical condition; or whether she takes a given medication); personal information is also an attribute of the respondent, but unlike private information the respondents do not normally mind that data collector knows the probability distribution of personal information (e.g., what is the probability that the color of the respondent's hair being black, or what is the probability that she lives near a lake). We also assume that private and personal information are unrelated - e.g., taking a medication is unrelated to one's hair color. To enhance the level of cooperation, instead of asking each respondent whether she has the attribute A, data collector asks each respondent two unrelated questions. One of them asks private information, i.e., the one that data collector is interested in. The other refers to personal information. The answers to the two questions are unrelated to each other [9]. For example, the survey questions can be designed as follows:

1. Do you have the *private* attribute A?
2. Do you have the *personal* attribute Y?

In practice, the first question could be "Are you taking medicine A?", and the second question could be "Do you live near a lake?". Respondents answer one of these two questions. They use a randomization device to decide which question to answer, without letting data collector know which question is answered. Each randomization device tells the respondent which question she is to answer: the probability of choosing

The first question is q , and the probability of choosing the second question is $1-q$. Although data collector learns a response (i.e., "yes" or "no"), he does not know which question was answered by the respondents. It is important to engineer the interaction between data collector and respondent in such a way that the respondent will trust the system, i.e., the respondent will clearly understand that data collector has no way of knowing which of the two questions is answered. Thus the respondent feels that her privacy is preserved. We further comment on this in Sec.4. Note that data collector only knows the probability distribution of the respondent's attribute Y. This is consistent with the interpretation of a personal attribute - data collector could know the distribution of the values (e.g., hair colors) of the personal attribute in the general population, without knowing the value of that attribute for a specific respondent.

The randomized response technique discussed above considers only one attribute. However, data sets usually consist of multiple attributes; finding the relationship among these attributes is one of the major goals for data mining. Therefore, we need techniques that can handle multiple attributes while supporting various data mining computations. In this paper, we provide multi-variant

randomized response technique (MRR) to address the problems of respondent privacy in electronic surveys.

2.1 Notations

In this work, we assume data are binary, but the techniques can be extended to categorical data. Suppose there are N *private* attributes (A_1, A_2, \dots, A_N) in a data set A. We construct N *personal* attributes (Y_1, Y_2, \dots, Y_N) . We want one *private* attribute (question) to pair with one *personal* attribute (question), therefore we make the number of attributes of Y and the number of attributes of A be equal. Let A and Y represent any logical expression based on those attributes $A_i (i \in [1, N])$ and $Y_i (i \in [1, N])$. For example, A can be $(A_1 = 0) \wedge (A_2 = 1)$ and Y can be $(Y_1 = 0) \wedge (Y_2 = 1)$. Let $P(Y)$ be the proportion of the records in the personal data that satisfy $Y = \text{true}$. Let $P^*(A)$ be the proportion of the records in the whole randomized data set that satisfies $A = \text{true}$. Let $P(A)$ be the proportion of the records in the whole non-randomized data set that satisfy $A = \text{true}$ (the potential non-randomized data set which in reality does not exist). $P^*(A)$ can be observed from the randomized data, but $P(A)$, the actual proportion that we are interested in, cannot be observed from the randomized data because the non-randomized data set is not available to anybody; we have to estimate $P(A)$. The goal of MRR is to find a way to estimate $P(A)$ from $P^*(A)$.

2.2 Multi-variant Randomized Response Scheme

In this scheme, all the attributes including the class label will be treated as a group. They either keep the same values or obtain the values from personal data. In other words, when sending the private data to data collector, respondents either tell their answers to the private questions or tell their answers to the personal questions. The probability for the first event is q , and the probability for the second event is $1-q$. For example, assume a respondent's attribute values A_1 and A_2 are 11 for private data; and the respondent's attribute values Y_1 and Y_2 are 01. The respondent generates a random number between 0 and 1; if the number is less than q she sends 11 to data collector; if the number is bigger than q , she sends 01 to data collector. Since data collector only knows q which is the same for all respondents and does not know the random number generated by each respondent, he cannot know whether the respondent tells the values from private data or personal data. To simplify our presentation, we use $P(A(11))$ to represent $P(A_1 = 1 \wedge A_2 = 1)$, $P(Y(11))$ to represent $P(Y_1 = 1 \wedge Y_2 = 1)$ where " \wedge " is the logical *and* operator. Because the contributions to $P^*(A(11))$ partially come from $P(A(11))$, and partially

come from $P(Y(11))$, we can derive the following equation:

$$P^*(A(11)) = P(A(11)) \cdot q + P(Y(11)) \cdot (1-q)$$

Since $P(Y(11))$ is known as Y is personal data, q is determined before collecting the data, and $P^*(A(11))$ can be directly computed on the disguised (randomized) data set. By solving the above equation, we can obtain $P(A(11))$, the information needed to build a naive Bayesian classifier. The general model is described in the following:

$$P^*(A) = P(A) \cdot q + P(Y) \cdot (1-q) \quad (1)$$

2.3 Building Naïve Bayesian Classifiers

The naive Bayesian classifier is one of the most successful algorithms in many classification domains. Despite of its simplicity, it is shown to be competitive with other complex approaches, especially in text categorization and content based filtering. The naive Bayesian classifier applies to learning tasks where each instance x is described by a conjunction of attribute values and where the target function $f(x)$ can take on any value from some finite set V . A set of training examples of the target function is provided, and a new instance is presented, described by the tuple of attribute values $\langle a_1, a_2, \dots, a_n \rangle$. The learner is asked to predict the target value for this new instance. Under a conditional independence assumption, i.e.,

$P(a_1, a_2, \dots, a_n | y_i) = \prod_{i=1}^n P(a_i | y_i)$, a naive Bayesian classifier can be derived as follows:

$$\begin{aligned} v_{NB} &= \operatorname{argmax}_{v_i \in V} P(v_i) \prod_{i=1}^n P(a_i | v_i) \\ &= \operatorname{argmax}_{v_i \in V} P(v_i) \prod_{i=1}^n \frac{P(a_i \wedge v_i)}{P(v_i)} \end{aligned}$$

To build a NB classifier, we need to compute $P(v_i)$ and $P(a_i \wedge v_i)$. To compute $P(v_i)$, we can use the general model (Eq.(1)) with A being $(C = v_i)$ and Y being $(CY = v_i)$ where C is the class label for the private data A and CY is the class label of personal data Y . $P^*(A)$ can be computed directly from the (whole) randomized data set. $P(Y)$ is known since it is personal and q is known as well. By knowing q , data collector, who conducts the training, only knows the probability of the training data being private, but does not exactly know if each value is private data or not. By solving the above equation, we can get $P(A)$ which is $P(C = v_i)$ in this case. Similarly, we can compute $P(a_i \wedge v_i)$ using the general model (Eq.(1)) with A being $(A_i = a_i \wedge C = v_i)$ and Y being $(Y_i = a_i \wedge CY = v_i)$.

2.4 Testing

Conducting the testing is straightforward when data are

not randomized, but it is a non-trivial task when the testing data set is randomized. When we choose a record from the testing data set, compute a predicted class label using the naive Bayesian classifier, and find out that the predicted label does not match the record's actual label, do we say this record fails the testing? If we knew whether the record represents the private or the personal data, and if we knew the true class for each data, we could easily answer this question. But how can we compute the accuracy score of a NB classifier when data are randomized? Our answer is to apply the multi-variant randomized response technique once again to compute the accuracy. Let us use an example to illustrate how to compute the accuracy. Assume the number of attributes is 2. To test a record $(A_1 = 1, A_2 = 0)$ denoted by $A(10)$, we feed $A(10)$ and $Y(10)$, where $Y = (Y_1 = 1, Y_2 = 0)$ to the NB classifier built in Sec.2.3. Let $P^*(A(cc))$ be the proportion of correct predictions using the disguised (randomized) testing data set, $P(Y(cc))$ be the proportion of correct predictions in the personal data, and let $P(A(cc))$ be the proportion of correct predictions in the private data. $P(A(cc))$ is what we want to estimate. Because $P^*(A(cc))$ consist of contributions from $P(A(cc))$ and $P(Y(cc))$, we have the following equation:

$$P^*(A(cc)) = P(A(cc)) \cdot q + P(Y(cc)) \cdot (1-q)$$

Where $P^*(A(cc))$ can be obtained from disguised testing data set. q is known and by knowing q , data collector, who conducts the testing, only knows the probability of the testing data being private, but does not exactly know if each value is private data or not. How does data collector know $P(Y(cc))$? One implementation is as follows: each respondent is given the same classifier by data collector. The classifier is constructed during the training (Sec.2.3). Each respondent applies this classifier on her personal data Y and communicates the number of correct predictions (0 or 1) to data collector, who then computes $P(Y(cc))$. Note that data collector does not know the values of the Y attributes, only the result of the classifier. Data collector can solve the above equation and get $P(A(cc))$, the accuracy score of testing.

3. EXPERIMENTAL RESULTS

To evaluate the effectiveness of our proposed scheme, We conducted experiments on two real life data sets *Adult* and *Breast Cancer* which were obtained from the UCI Machine Learning Repository.

3.1 Experimental Steps

We modified naive Bayesian classification algorithm to handle randomized data based on our proposed scheme. We applied our scheme to obtain a privacy-oriented classifier. We also ran naive Bayesian classification algorithm on original data set, and obtained a base classifier. We then applied the same testing data to both

classifiers. Our goal is to compare classification accuracy of these two classifiers. Obviously we want accuracy of privacy-oriented classifier to be close to accuracy of the base classifier. Our experiments consist of the following steps:

Step I: Preprocessing

Since we assume that data set contains only binary data, we first discretize original non-binary data to become binary. We split the value of each attribute from the median point of the range of the attribute. After preprocessing, we randomly divided data sets into a training data set D (80%) and a testing data set B (20%). Note that B will be used for comparing our results with benchmark results.

Step II: Benchmark

We use D and the original NB classification algorithm to build a classifier T_D ; we use data set B to test the classifier, and get an accuracy score. We call this score original accuracy (or benchmark score).

Step III: q Selection

For $q = 0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ and 1.0, we conduct the following 4 steps:

Sub-step 1: Randomization

We create a disguised data set G. For each record in training data set D, we generate a random number r from 0 to 1 using uniform distribution. If $r < q$, we copy the record of D to G without any change; if $r \geq q$, we randomly generate the values for a record of Y according to the pre-defined probability and copy the record values to G. In this paper, each record of Y is randomly generated such that each logical expression (Y) appears with the probability of 0.5. That is $W_y = 0.5$ (ref.Sec.4). We perform this randomization step for all the records in the training data set D, then generate the new data set G.

Sub-step 2: Classifier Construction

We use data set G and our modified NB classification algorithm to build a naive Bayesian classifier T_G .

Sub-step 3: Testing

We use data set B to test and get an accuracy score S.

Sub-step 4: Repeating

We repeat steps 1-3 for 1000 times, and get $S_1, S_2, \dots, S_{1000}$. We then compute mean and variance of these 1000 accuracy scores.

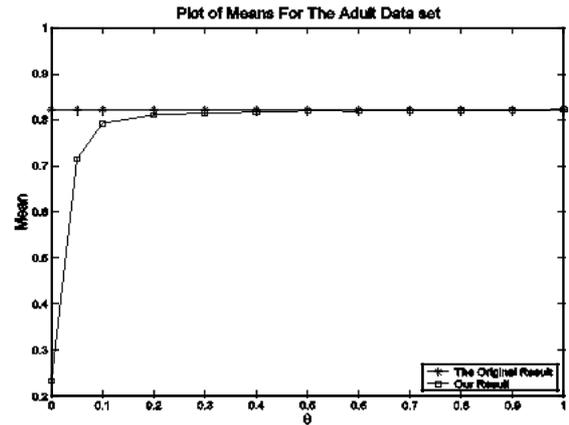


Figure 1

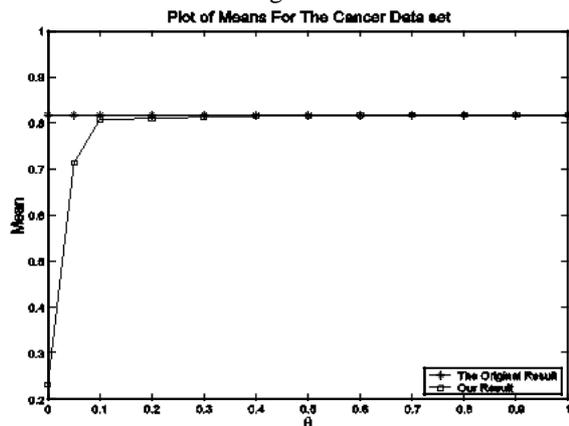


Figure 2

3.2 Accuracy Analysis

3.2.1 Analysis of Mean

Fig.1 and Fig. 2 shows mean values of accuracy scores for *Adult* and *Breast-Cancer* data sets respectively. We can see from figures that when $q=1$, the results are exactly the same as the results when standard, unmodified classification algorithm is applied. This is because when $q=1$, randomized data sets are all from private data D. When q approaches 1, contribution of private data is enhanced; with q deviating from 1, the contribution of private data is decreasing (when $q=0$, collected data set is all from personal data). Therefore, when q moves from 1 towards 0, the mean of accuracy has the trend of decreasing.

3.2.2 Analysis of Variance

Fig. 3 and Fig.4 shows variances of accuracy scores. When q moves from 1 towards 0, the degree of randomness in disguised data is increasing, variance of estimation used in our method becomes larger. Variance changes with different randomization levels q . When q is near 0, randomization level is much higher and private data is better disguised. We do not show variance when $q=0$. In this case, since collected data set is actually personal data and the probability

distribution for it is always the same for each iteration, variance is 0.

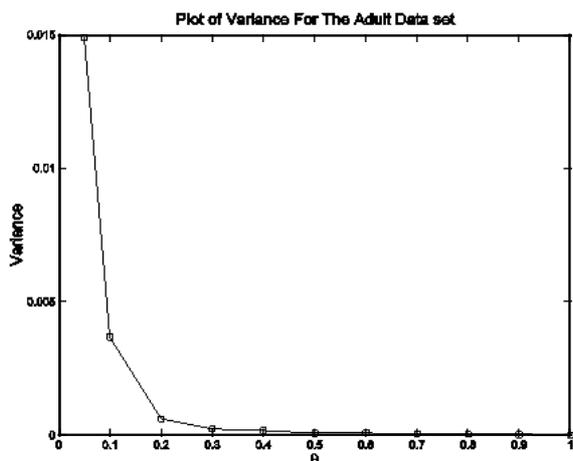


Figure 3

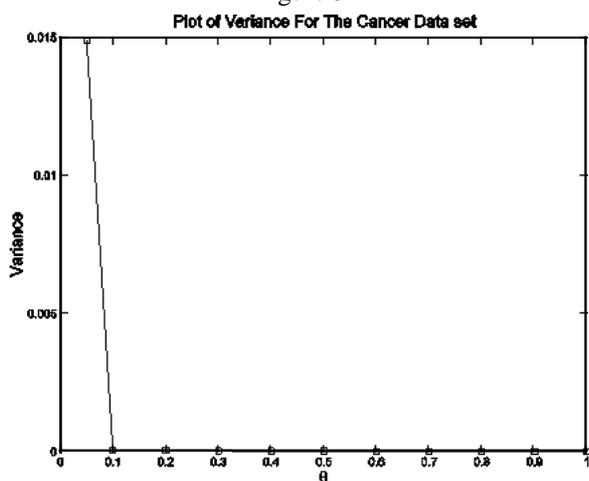


Figure 4

4. MEASURING PRIVACY

Various privacy enhancing techniques [5, 8, 11, 14, 15, 16] have been developed to protect data privacy. To trust a privacy enhancing technology, we need to know how much privacy that a particular technique protects. A challenge faced by privacy-preserving data mining community is how to measure privacy. A general privacy measure which can quantify privacy for any privacy protection scheme is still an open question. In this section, we develop a privacy measure for our proposed multi-variant randomized response technique. Our measure contains two steps: First, we measure privacy for a single entry. Second, we select the minimal privacy value and treat it as the privacy level for the group. The reason why we choose the minimal value for the group is that, when the entries are randomized together, finding the original value for one entry will cause disclosing the original values for other entries in the group.

For a single entry, original value can be 1 or 0; randomized value can be 1 or 0 as well. Privacy comes from uncertainty of original value given a randomized

value. In other words, if original value is 1, given randomized value 1 or 0, privacy will be the probability that data collector guesses the original value being 0. There are four possible randomization results:

- Original value is 1, value after randomization is 1;
- Original value is 1 but value after randomization is 0;
- Original value is 0 but value after randomization is 1;
- Original value is 0, value after randomization is 0.

Let's use the following denotations:

- Let's O_m be the original value;
- Let's R_m be the value after randomization;
- Let's G_m be the guessed value.
- Let's W_a be the probability that a value is 1 in data set A, and the probability that a value is 0 in data set A is $(1-W_a)$;
- Let's W_y be the probability that a value is 1 in data set Y, and the probability that a value is 0 in data set Y is $(1-W_y)$;

Privacy for a single entry, denoted by PSE , can be derived as follows:

$$\begin{aligned}
 PSE = & \Pr(O_m = 1) \cdot \Pr(R_m = 1 | O_m = 1) \cdot \Pr(G_m = 0 | R_m = 1) \\
 & + \\
 & \Pr(O_m = 1) \cdot \Pr(R_m = 0 | O_m = 1) \cdot \Pr(G_m = 0 | R_m = 0) \\
 & + \\
 & \Pr(O_m = 0) \cdot \Pr(R_m = 1 | O_m = 0) \cdot \Pr(G_m = 1 | R_m = 1) \\
 & + \\
 & \Pr(O_m = 0) \cdot \Pr(R_m = 0 | O_m = 0) \cdot \Pr(G_m = 1 | R_m = 0) \\
 = & Comp_1 + Comp_2 + Comp_3 + Comp_4
 \end{aligned}$$

The first component contains three parts:

1. $\Pr(O_m = 1)$ is the probability that a value is 1 in private data set (A), which is W_a .
2. $\Pr(R_m = 1 | O_m = 1)$ is the probability that a randomized value is 1 given the original value is 1. There are two possibilities: (1) a randomized value comes from data set A, and the probability for the case is q ; (2) a randomized value comes from data set Y, and the probability for this case is $(1-q) \cdot W_y$.
3. Let $(G_m | R_m)$ be the guessed value given a randomized value and $(O_m | R_m)$ be the original value given the same randomized value. For the same R_m value, there are two possibilities: $G_m = O_m$ or $G_m \neq O_m$. We should notice when $(G_m | R_m) \neq (O_m | R_m)$, there is zero contribution to PSE . For instance, in the first component, the original value is 1, the guessed value is 0 given the randomized value is 1. When $(G_m | R_m) \neq (O_m | R_m)$, then the guessed will be 1. That is the guessed value and the original value is the same (both of them are 1s) and it contributes zero to the PSE . Therefore, we only consider the case where

$(G_m | R_m) = (O_m | R_m)$ for the third part. We can apply Bayes rule to tackle this part. After applying the Bayes rule, we obtain $\Pr(R_m=1 | O_m=0) \cdot \Pr(O_m=0) / \Pr(R_m=1)$. $\Pr(O_m=0)$ is the probability that a value is 0 in private data set (A), which is $1 - W_a$. $\Pr(R_m=1 | O_m=0)$ is the probability that a randomized value is 1 given the original value is 0. In this case, randomized value cannot come from private data since original value is 0 and randomized value is 1. The only possibility is that randomized value is from personal data set Y, and the probability is $(1-q) \cdot (1 - W_y)$. As for $\Pr(R_m=1)$, we can extend this term and details are shown as follow:

$$\begin{aligned} Comp_1 &= W_a \cdot [q + (1-q) \cdot W_y] \cdot \frac{\Pr(R_m=1 | O_m=0) \cdot \Pr(O_m=0)}{\Pr(R_m=1)} \\ &= \frac{W_a \cdot [q + (1-q) \cdot W_y] \cdot (1-q) \cdot (1 - W_y) \cdot (1 - W_a)}{\Pr(R_m=1 | O_m=1) \cdot \Pr(O_m=1) + \Pr(R_m=1 | O_m=0) \cdot \Pr(O_m=0)} \\ &= \frac{W_a \cdot [q + (1-q) \cdot W_y] \cdot (1-q) \cdot (1 - W_y) \cdot (1 - W_a)}{[q + (1-q) \cdot W_y] \cdot W_a + (1-q) \cdot (1 - W_y) \cdot (1 - W_a)} \end{aligned}$$

Similarly, we can obtain other components, and we then get

$$\begin{aligned} PSE &= \frac{(1 - W_a) \cdot W_a \cdot (1-q) \cdot [q + (1-q) \cdot W_y]}{[q + (1-q) \cdot W_y] \cdot W_a + (1-q) \cdot W_y \cdot (1 - W_a)} \\ &\quad + \\ &\quad \frac{2W_a \cdot (1 - W_a) \cdot (1-q) \cdot (1 - W_y) \cdot [q + (1-q) \cdot (1 - W_y)]}{[q + (1-q) \cdot (1 - W_y)] \cdot (1 - W_a) + (1-q) \cdot (1 - W_y) \cdot W_a} \end{aligned} \quad \text{Eq. (2)}$$

We compute *PSE* for each single entry. We then select the smallest value *PSE*(Min) as the privacy value for the group. We can see from Eq.(2) that, *PSE* is determined by three parameters: (1) control parameter *q*; (2) private data (ref. to as data set A in Sec.2) distribution W_a ; (3) personal data (ref. to as data set Y in Sec.2) distribution W_y . What we can see from *PSE* equation is that, when $W_a = 0.9, 0.8, 0.7, 0.6$ privacy is symmetric with respect to privacy when $W_a = 0.1, 0.2, 0.3, 0.4$. In other words, given a certain *q*, *PSE* value will equal to the *PSE* value when control parameter is $1 - q$.

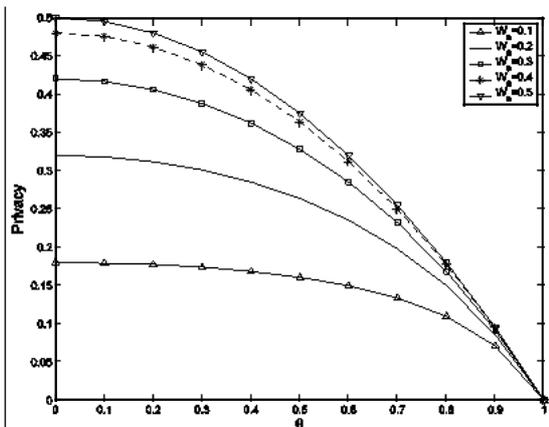


Figure 5

To get better sense of our proposed privacy measure, we conducted a set of experiments on private data sets with

various distributions and personal data set with $W_y = 0.5$. Specifically, we conduct experiments when $W_a = 0.1, 0.2, 0.3, 0.4, \text{ and } 0.5$. For each data distribution, we compute privacy value for the cases where $q = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$. As we see from results in Fig. 5:

- When $q=1$, private data is fully disclosed. Privacy value is 0;
- When $q=0$, data collector gets no private data, and the data obtained are all personal data. In this case, privacy level of private data is the highest.
- When q is away from 1 and approaches 0, the elements of private data contribute less to the classification, and the probability of disclosing private data is decreasing. Therefore privacy level of private data increases.
- When private data (A) distribution approaches to uniform ($W_a = 0.5$), privacy level is increasing. Since uniform distribution will make original data recoverability be the lowest.

Empirical results from Sec.3 and Sec.4 confirm that recoverability and privacy are complementary goals. Given W_a and W_y , the best privacy is achieved when control parameter q is 0; however, the accuracy will be the worst in this case. The best accuracy is attained when $q=1$ but privacy is the worst. Trade-offs are also applied when q has a value between 0 and 1. In practice, how to select q is dependent upon our primary goals. If we want to the results be very precise, we need choose the values near 1; in contrast, if privacy is the primary goal, we choose the values near 0.

6. CONCLUDING REMARKS

In this paper, we have presented a method to build naïve Bayesian classifiers using multi-variant randomized response technique. Experimental results show that when we select an appropriate randomization parameter q , we can get fairly accurate classifiers comparing to the classifiers built from undisguised data. A privacy measure was developed and privacy analysis was also conducted. Trade-offs between privacy and accuracy are discussed. The proposed multi-variant unrelated question model has a broader impact in the sense that it can be used not only for naïve Bayesian classification, but also can be utilized in many other privacy-preserving data mining computations, such as decision tree induction, Bayesian classification, probabilistic-based clustering. As future work, we will apply the proposed scheme to other data mining problems.

REFERENCES

[1] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "Random Data Perturbation Techniques and Privacy Preserving Data Mining", the IEEE International

- Conference on Data Mining, 2003, Florida, USA.
- [2] A. Evfimievski, J. E. Gehrke, and R. Srikant, "Limiting Privacy Breaches in Privacy Preserving Data Mining", Proceedings of the 22nd ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, 2003, San Diego, CA.
- [3] R. Agrawal and R. Srikant, "Privacy-Preserving Data Mining", Proceedings of The ACM SIGMOD Conference On Management of Data, 2000, Dallas, Texas, USA.
- [4] L. F. Cranor and J. Reagle and M. S. Ackerman, "Beyond concern: Understanding net users' attitudes about online privacy", AT&T Labs-Research, 1999, April, Available from <http://www.research.att.com/library/trs/TRs/99/99.4.3/report.htm>
- [5] W. Du and Z. Zhan, "Using Randomized Response Techniques For Privacy-Preserving Data Mining", Proceedings of The 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003, Washington, DC, USA, August 24-27
- [6] Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining", Advances in Cryptology - CRYPTO '00, 2000, 1880 of Lecture Notes in Computer Science. Springer-Verlag, 36-54
- [7] Langley, P., Iba, W., and Thompson, K., "An Analysis of Bayesian Classifiers, National Conference on Artificial Intelligence, 223-228, 1992, url = "citeseer.nj.nec.com/langley92analysis.html"
- [8] S. Rizvi and J.R. Haritsa, "Maintaining Data Privacy in Association Rule Mining", Proceedings of the 28th VLDB Conference, 2002, Hong Kong, China
- [9] S. Warner, "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias", The American Statistical Association, Volume:60, number:309, pages:63-69, March,1965
- [10] J. Vaidya and C. Clifton, "Privacy-Preserving K-Means Clustering over Vertically Partitioned Data", Proceedings of The 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003, Washington, DC, USA

Proposal of Improving Model for Default Probability Prediction with Logit Model on Non-Compensatory Rule

Sakamaki, Yoshikazu

Department of Value and Decision Science, Tokyo Institute of Technology, Tokyo, 1528550, Japan

ABSTRACT

Financial statement analysis is widely used for credit risk analysis. This method was developed at the end of 19's for the purpose of surveying credit reliability of credit customers. However, the result of analysis with financial statements is likely to be controlled by the work experience of analysts. As the result, it is difficult to maintain consistency of risk measurement, and moreover, it is expensive and time consuming to perform the vast amount of precise evaluations required in a short time. Many researchers have studied financial statement analysis scientifically applying mathematical and statistical methods, especially with use of multivariate statistical analysis.

This report presents the history of preceding risk control studies and considers issues of the preceding default forecasting model based mainly on the binominal logit model on compensatory rule. Moreover we propose the improving scoring model on the non-compensatory rule and verify that our model is superior to the binominal logit model, and is useful in business using based on actual financial data.

1. INTRODUCTION

During the period of long economic stagnation, Japan has been faced with the task of reestablishing the reliability of its financial systems. The financial markets must be reformed in order to obtain high evaluations worldwide. The Financial Services Agency claims that they will reduce the bad debt ratio of Japanese main banks by 50% by the end of 2004, and resolve the issue of non-performing loans. Adding to this policy, they also suggest that they intend to build a much stronger financial system that can support the promotion of structural reform. To achieve our goal, all financial institutes, personal investors and business corporations must recognize the existence of credit risk and understand the techniques of risk control.

Financial statement analysis is conventionally used as the basic method of credit risk analysis. Credit risk analysis based on financial statements developed at the end of the 19's for the purpose of surveying the credit reliability of loan customers. However, the result of financial statement analysis is likely to be controlled by the level of work experience of analysts. As a result, it is difficult to maintain the consistency of risk measurement on such a subjective basis. Moreover, it is expensive and time consuming to perform the vast amount of precise risk analysis in the limited time.

Since W. H. Beaver (1967)[1], many researchers have studied financial statements analysis scientifically using mathematical and statistical methods, especially with the use of multivariate statistical analysis. One of the most well known studies is the Z-score model proposed by E. I. Altman(1966)(1968).

This report presents the history of the statistical approach of risk control and considers issues of the preceding default forecasting model with the linearly linked binominal logit model on compensatory rule, which is

conventionally used as the scoring model. In addition, an improved version of the scoring model with the binominal logit model on non-compensatory rule is presented.

2. PRECEDING STUDY

This chapter introduces studies of credit risk measurement with a statistical approach in the default forecasting model. As mentioned in Chapter 1, the study of default forecasting originates from Altman's discriminant model (1968). The result of his study is well known as the Z-score model, and it remains one of the most widely used methods for default forecasting all over the world.

Altman's z-score model is based on discriminant analysis. In this model, they assume that each firm has its own risk factor, assigning appropriate weight to each risk factor and adding all of the weighted factors. As the result, we obtain a linearly linked risk factor for a bankrupt company and a non-bankrupt company.

That is, let x_{mi} be the m^{th} risk factor for company i ($m=1,2,\dots,M$) and β_m be the parameter for variable x_{mi} , and we define z-score as

$$z_i = \sum_{m=1}^M \beta_m x_{mi} \quad (1).$$

Here, the company is default if $z_i < 0$, and non-default if $z_i > 0$.

Now, we define

$Z^{(1)}$: Average of z-score for default companies

$Z^{(2)}$: Average of z-score for non-default companies

---(2)

and let variance of the linearly-linked risk factors through all the sample data be $Var(Z)$. Then, unknown parameters are given by maximizing

$$\eta^2 = \frac{(\bar{z}^{(1)} - \bar{z}^{(2)})^2}{\text{Var}(z)} \quad (3)$$

With the discriminant score calculated from the linearly linked risk factors weighted by the estimated parameters, we can estimate whether each debtor belongs to the default or non-default group.

The evaluation methods of credit risk based on the linear discriminant function is mathematically easy to handle and analogous to traditional credit risk evaluation analysis using a score card. That is the reason why the scoring model with discriminant analysis is used all over the world. However, when we apply discriminant analysis to credit risk control, we have to be very careful of the following statistical assumptions specific for credit risk measurement.

<1> Normality and variance equivalence

If the variables used in the discriminant analysis is not distributed normally or variance is not equivalent between the default and the non-default group, default forecasting with the discriminant score and statistical testing for the parameters cannot always be performed correctly.

<2> Uniformity of default/non-default probability

When we use discriminant analysis, we cannot have any information about the ratio of the default and the non-default firms. The discriminant model is usually built with the assumption that the ratio of the default firms and the non-default firms is equal. However, the ratio of default/non-default firms generally differs among banks.

<3> Meaning of calculated weighting parameters

Unlike the parameters estimated in regression analysis, the absolute value of discriminant parameters is not decided uniquely, and only the relative ratio among parameters is decided uniquely. That is, the absolute score in discriminant analysis cannot also be decided uniquely. In discriminant analysis, only the distance from the discriminant point is statistically significant.

Discriminant analysis is widely used but, as shown above, it involves the issue of statistical difficulty. Moreover, we need to make clear when bankruptcy occurs and what the default probability is, in order to qualify the credit risk. Even if we can know the relative default score as the result of discriminant analysis (z-score), we cannot know the absolute default probability.

To solve such an issues, normal regression analysis has begun to be used. Next we consider linear regression analysis. That is, let

y_i : Bernoulli random variable (If firm i is non-default, it takes 0. And if firm i is defaulted, it takes 1.) ($i=1,2,\dots,J$)

x_{ij} : j^{th} credit risk factor of firm i ($j=1,2,\dots,J$)

β_j : Estimated parameter for j^{th} risk factor

and we define the default probability of firm i by

$$y_i = \beta_0 + \sum_{j=1}^J \beta_j x_{ij} \quad (4)$$

And the expected value of the dependent variable given by equation (4) becomes the estimated default probability. However, in the linear regression analysis, when the summation of all of the weighted risk factor becomes very large, the default probability exceeds 1. On the other hand, if the summation becomes very small, the default probability is less than 0. And it becomes contradictory to the definition of statistics. To solve this issue, the following method is used to adjust the expected default probability estimated through the above regression analysis. That is, when the expected default probability exceeds 1, we truncate the default probability to 1, and when the expected default probability is less than 0, we truncate the default probability to 0. However, it is not statistically natural to put the expected default probability between 0 and 1 compulsorily.

To resolve these issues, logistic regression analysis (binominal logit model on compensatory rule) began to be used (Martin;1977)[3]. In the logistic regression analysis, let the summation of all of the weighted risk factors be

$$Z_i = b_0 + b_1 x_{i1} + \dots + b_m x_{im} \quad (5)$$

and the default probability of firm i (p_i) be

$$p_i = \frac{1}{1 + \exp(-Z_i)} \quad (6)$$

If the default probability of firm i is given by equation (6), the likelihood function for parameter vector \mathbf{B} is given by

$$L(\mathbf{B}) = \prod_{i=1}^I p_i^{y_i} (1 - p_i)^{1-y_i} \quad (7)$$

Here, $y_i=1$: If firm i is defaulted and $y_i=0$ if firm i is not defaulted

Moreover, log-likelihood function is defined by

$$l(\mathbf{B}) = \sum_{i=1}^I \{y_i \log p_i + (1 - y_i) \log(1 - p_i)\} \quad (8)$$

Generally, parameters that maximizes equation (8) are estimated by the maximum likelihood estimation method.

3. MODEL IMPROVEMENT

3-1. Issues in the preceding model

Default probability estimate based on logistic regression analysis is very significant because it enables measurement of default probability on an absolute scale, which cannot be performed in the classical credit risk models such as that of Altman (1968). However, in the normal logit model based on the compensatory rule, even if only one risk factor deteriorates, the default probability becomes very high although the firm does not actually confront management distress.

Regarding this point, researchers have shown very interesting report in preceding studies of marketing

research. In the study of marketing research, compensatory rule is used widely in the consumer's choice behavior model today because compensatory rule is widely believed to be able to mimic consumer's choice behavior on non-compensatory rule, when the following two conditions are satisfied.

That is,

- (1) Attributes are related monotonically to consumer's preferences.
- (2) There is no error or uncertainty about these preferences. (Dawes and Corrigan (1974)[4]).

Because these conditions are thought likely to be satisfied in most choice behavior cases, compensatory models are widely used even when it does not reflect the actual decision making process (Cattin and Wittink (1982)[5]).

Despite these reports, there is still reason for concern that compensatory models may not always mimic non-compensatory processes. Several researchers (Curry Louviere, and Augustine (1981)[6]; Einhorn, Kleinmuntz and Kleinmuntz (1979)[7]) have argued that studies demonstrating robustness in linear models have underemphasized a major influence on predictive accuracy: the correlation among attributes of each alternatives.

Newman (1977) was the first to examine explicitly the relationship of correlations among attributes and the fit of linear models, following brief discussions offered by Einhorn (1970)[8] and Goldberg (1971) [9]. He noted that though Wainer et al (1976)[10] conjecture that small departures from optimal coefficients in a linear model do not make no nevermind, will often hold in practice, it will not hold when the attributes describing alternatives are negatively correlated.

Negative correlations also can diminish model performance because they make predictions more sensitive to the particular choice strategy used Einhorn, Kleinmuntz, and Kleinmuntz (1979) illustrate this effect through the example of a negatively correlated environment containing two options, described on two attributes.

The first is good on attribute A and bad on B whereas the second is bad on A and good on B. The option that will be picked in this case depends on how a choice strategy weights the two attributes. A lexicographic strategy that selects the option best on attribute A, for example, would choose the first option. A similar lexicographic strategy looking first at attribute B would pick the second. In contrast, a compensatory rule that weights the two attributes equally would be indifferent. If the two attributes were correlated positively, if the first option were good on both attributes and the second bad on both, one would pick the first option regardless of how the attributes are weighted. Hence, in general, the less the redundancy among attributes in a choice environment

(the more negative the average inter-attribute correlation), the greater the difference in predictions made by different models.

A more precise discussion regarding this issue was made by Curry and Faulds (1986), who described the theoretical relationship that should be present between two different linear weighting schemes given different correlational structures. They noted that when two models are compared in a maximally negative inter-correlational environment, even small deviations in their weight can induce major discrepancies in prediction, approaching a complete rank-order reversal in overall preferences for options.

In this study, we propose an improving default probability forecasting model based on these discussions.

Table 1 shows the basic statistical value and correlation matrix of trunk seven representative financial ratios of default firms and non-default firms on the BULK database system provided by Nikkei group. As you can see, in the firms confronting bankruptcy, most of the financial indexes deteriorate compared with non-default firms.

Table.1 Basic statistics and correlation matrix

	Xa	Xb	Xc	Xd	Xe	Xf	Xg	
Xa	Capital Adequacy Ratio	1	-0.80402 <.0001	-0.37415 <.0001	0.54872 <.0001	0.40042 <.0001	0.16345 <.0001	0.4203 <.0001
Xb	Ratio of Check account degree of dependence		1	0.47522 <.0001	-0.41472 <.0001	-0.41532 <.0001	0.02308 <.0001	-0.25431 <.0001
Xc	Ratio of interest expenses to sales			1	-0.3817 <.0001	-0.24367 <.0001	0.42748 <.0001	-0.17885 <.0001
Xd	Acid Ratio				1	0.44421 <.0001	0.06875 <.0001	0.71177 <.0001
Xe	Deposit-Loan Ratio					1	0.14732 <.0003	0.57603 <.0001
Xf	Ratio of Return on Sales						1	0.17983 <.0001
Xg	Cash Ratio							1

Non-Default	N	average	std	minimum	maximum
Capital adequacy ratio	2820	0.3296	6.5237	-385.5433	0.9994
Ratio of check account degree of dependence	2820	0.3646	4.4858	0.0000	265.6260
Ratio of interest expenses to sales	2820	232.2219	13766.2900	0.0000	816400.0000
Acid Ratio	2820	1.2452	2.3485	0.0000	77.6429
Deposit-loan ratio	2820	24.0607	523.9959	-0.0214	19008.5000
Ratio of Return on Sales	2820	-127.0403	7528.9700	-446800.0000	0.9042
Cash Ratio	2820	0.5765	2.2616	-0.0069	77.6429
Default	N	average	std	minimum	maximum
Capital adequacy ratio	94	-0.4427	1.7729	-8.8748	0.8520
Ratio of check account degree of dependence	94	1.1663	2.0433	0.0000	16.4592
Ratio of interest expenses to sales	94	0.0415	0.1477	0.0000	1.4352
Acid Ratio	94	0.4592	0.5277	0.0087	3.0995
Deposit-loan ratio	94	0.0999	0.1380	-0.0271	1.0553
Ratio of Return on Sales	94	-0.5719	5.1200	-49.6661	0.4249
Cash Ratio	94	0.1342	0.2428	-0.0564	2.0455

*Correlation is calculated with use both of default and non-default data.

As shown in Table 1, on many risk factors, especially firms confronting bankrupt have deteriorating financial ratios. That is, it can be predicted that all of the financial ratios should become worse at the same time when bankruptcy occurs. And there seems to be very negatively strong correlation among the major financial ratios.

Taking these points into consideration, applying the logit model on the non-compensatory rule in which default probability increases only when all of the risk factors used in the model deteriorate seems to be more appropriate than applying the model of the compensatory rule.

In this study, we propose an improving default probability forecasting model based on non-compensatory rule. Moreover, we apply the actual financial data to our model and reveal a higher default forecasting power than the preceding model.

3-2 Model Improvement

In this section, we show two ways of improving the default probability forecasting model: the conjunctive type and the disjunctive type.

<Conjunctive type>

In this type of model, borderline for each risk factor is decided, and we assume that firm fails only when all risk factors go over the borderline. Borderline varies among risk factors, and if we assume that correlation among error terms become zero implicitly and variation fluctuates based on logistic distribution at random, we can define the probability for firm i to fail as

$$\begin{aligned} P_i(y_i = 1) &= P_i(x_{ik} - t_k + e_{ik} \geq 0, \forall k) \\ &= \prod_{k=1}^K P_i(x_{ik} - t_k + e_{ik} \geq 0) \\ &= \prod_{k=1}^K \frac{1}{1 + \exp(-b_k(x_{ik} - t_k))} \end{aligned} \quad (9)$$

It means that the default probability for firm i is in proportion to the distance from the borderline of each risk factor.

<Disjunctive type>

In this type of model, we assume that firm fails if, at least, any one of the risk factors goes over the borderline. Just as mentioned in the conjunctive type model, assuming that borderline varies among firms and variation fluctuates based on logistic distribution, we can derive the following equation.

$$\begin{aligned} P_i(y_i = 1) &= P_i(\exists k \ x_{ik} - t_k + e_{ik} \geq 0) \\ &= 1 - P_i(x_{ik} - t_k + e_{ik} < 0, \forall k) \\ &= 1 - \prod_{k=1}^K P_i(x_{ik} - t_k + e_{ik} < 0) \\ &= 1 - \prod_{k=1}^K \left(1 - \frac{1}{1 + \exp(-b_k(x_{ik} - t_k))}\right) \end{aligned} \quad (10)$$

Here,

x_{ik} : k^{th} risk factor of firm i ($i=1,2,\dots,I; k=1,2,\dots,K$)

β_k : Scale parameter for k^{th} risk factor ($k=1,2,\dots,K$)

t_k : Border line of k^{th} risk factor ($k=1,2,\dots,K$)

y_i : Bernoulli random variable (If firm i is non-default, it takes 0. And if firm i is defaulted, it takes 1.)

3-3 Comparison model

As the comparison model, we use the normal logit model shown by equation (6).

3-4 Parameter estimation

Parameters are estimated by equation (11) based on the MLE (Maximum Likelihood Estimation method). If we use MLE for the parameter estimation of nonlinear model, parameter solution sometimes drops in the local

minimum.

To avoid this issue, parameters are estimated by comparing all of the combination patterns of variables to decide the most suitable variable combination, and we generated 20 combination patterns of values at random in advance which was used as the initial value for each combination pattern of variables.

$$L(b) = \prod_{i=1}^I p_i^{y_i} (1 - p_i)^{1 - y_i} \quad (11)$$

3-5 Calibration of model fitness

In this study, log-likelihood, AIC criteria, Kolmogorov-Smirnov distance, and divergence are used as the index of model fitness to compare the appropriateness of the normal linearly linked binomial logit model on the compensatory rule and improving logit model on the non-compensatory rule.

Here, the Kolmogorov-Smirnov test measures the maximum distance between two distribution functions. In our case, we look at the difference between the default/non-default distribution. This difference is known as the KS distance and it is used to assess the fitness degree of our model.

Divergence is the statistical index that shows the separation of the distribution of the default and non-default firm, which is calculated with the average and variance of the score.

It can be calculated by

$$Divergence = \frac{2(m_A - m_B)^2}{V_A + V_B} \quad (12)$$

Where,

μ_A, μ_B : Expected value of score given to default/non-default firms

V_A, V_B : Variance of score given to default/non-default firms

4. CALIBRATION

This chapter verifies the appropriateness of the proposed model. Publicly disclosed financial data of companies sold by Nikkei Group (BULK system) is used as the calibration data.

4-1 Outline of calibration data

1. Data source: Account settlement data of the company (From April 2000 to March 2001)

2. Data volume: 2,914 companies

3. Default ratio:

If the company defaulted in three years after the accounting date, the company is regarded as defaulted. (Default ratio: $84/2,914=2.88\%$)

Here, especially, in the case of normal logit model, which has a linear structure of risk factors, outliers in financial ratios often cause the problem of biased default probability estimation. In this study, we set the lower cap at the 1 percentile point and the upper cap at 99 percentile point.

4-2 Parameter estimation

Based on this actual financial data, we estimate model parameters for the normal logit model, conjunctive /disjunctive type of logit model on the non-compensatory rule.

Generally, the fitness of the model on the training data set is fine, but the most important question is how the

model works in out-of-sample data. So that in estimating parameters, we have chosen 60% of the sample data at random and used for in-sample-data, and 40% for out-of-sample data. Moreover, we generated 20 combination patterns of values at random in advance and used as the initial value of the maximum likelihood estimation method for the search algorithm, and prevented from dropping to the local minimum. Based on the value of the Log-Likelihood and AIC criteria, the proposed model adopts the estimated parameters which have minimum Log-Likelihood and AIC as the optimum parameters. It is probable that we obtain the result by chance, so that we tested several datasets to examine the superiority of our model. We show one of the results of parameter estimates in Table 3 due to the limitation of spaces.

Table 2 Result of parameter estimation

Parameters	Normal Logit Model		Compensatory Model (Conjunctive)		Non-Compensatory Model (Disjunctive)	
		P-value		P-value		P-value
B_0 Intercept	0.0021	0.00001	-	-	-	-
B_a Capital adequacy ratio	-3.9091	0.0219	-8.21248	0.00036	-2.94515	0.01859
t_a	-	-	0.21756	0.00083	0.10701	0.02643
B_b Ratio of check account degree of dependence	-0.63388	0.01226	5.51058	0.01922	2.48401	0.00053
t_b	-	-	0.56535	0.01228	0.13795	0.09182
B_c Ratio of interest expenses to sales	-10.9863	0.51632	-	-	-	-
t_c	-	-	-	-	-	-
B_d Acid Ratio	-1.93705	0.03417	4.76147	0.0274	6.9669	0.05109
t_d	-	-	-3.63163	0.0917	-2.07471	0.07811
B_e Deposit-loan ratio	-5.31485	0.03125	-	-	-	-
t_e	-	-	-	-	-	-
B_f Ratio of return on sales	-3.59059	0.00069	-1.85393	0.13607	-	-
t_f	-	-	-0.01782	0.0021	-	-
B_g Cash Ratio	-	-	-	-	-	-
t_g	-	-	-	-	-	-
Log-Likelihood(-2logL)	357.25248	-	310.964	-	386.42998	-
AIC	369.25248	-	326.964	-	398.42998	-
Divergence	0.153	-	0.162	-	0.108	-
K-S Distance	71.66%	-	75.26%	-	69.69%	-

Here, parameter estimates are obtained from in-sample data. And Log-Likelihood (-2logL), AIC, Divergence and K-S Distance are calculated based on the out-of-sample data.

5. RESULTS

5-1 Findings

As we can see in the financial policy for the revision of capital adequacy ratio based on BASEL II, application of rating technology for credit risk control is attracting significant attention from all over the world. For example, in BASEL II, it is permitted for each bank to use an original rating model, so-called internal rating grades, to grasp the total credit risk. And to realize the suitable increase of capital adequacy ratio, a more elaborate statistical model that can grasp total credit risk in the bank on time is needed. The linearly linked binominal logit model on the compensatory rule is used as the conventional model. However, generally, financial ratios that have strongly negative correlation are often used as a risk factor in the model, and as mentioned in the previous chapter, when we apply the normal logit model to the risk factors that have mutually negative correlations, it is expected that there exist stochastic issues.

So, this study proposed a default probability forecasting

model on the non-compensatory rule and resolved these issues. Simultaneously, based on the actual financial ratio data, we have shown that our model was superior to the preceding model.

5-2 Considerations

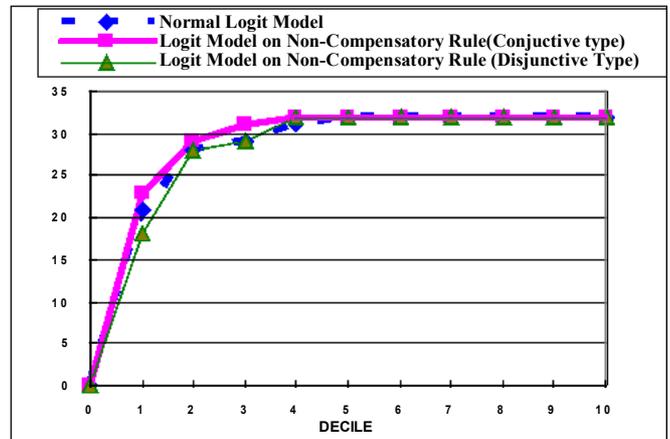


Figure 1 Comparison of Lift Curve

Figure 1 shows the lift curve of the number of bankrupt firms based on both of the linearly linked normal logit model on the compensatory rule and the improving logit model on the non-compensatory rule. As shown in this graph, the logit model on the non-compensatory rule (disjunctive type) does not have so high default capturing power, but the improved logit model on the compensatory rule (conjunctive type) shows a much higher default capturing power than the linearly linked normal logit model on the compensatory rule. It can also be proved by log-likelihood and the AIC statistical values shown in Table 3. As a result, we can conclude that the non-compensatory rule with conjunctive type has significantly superior power for default forecasting.

As mentioned in the previous chapter, when there exist strongly negative correlations among variables of the model, it is suggested that model performance reduces on the model of compensatory rule like normal logit model.

At first, to confirm the level of correlation among some financial indexes, we made correlation matrix and confirmed that there existed strongly negative correlations among some variables. (Show Table 1).

And after that, we proposed improving default probability forecasting model that enabled to improve the model performance. We proposed two types of models (Conjunctive type and disjunctive type), and as the result of model validation, we confirmed that we could improve model performance dramatically with use of conjunctive type model.

It can be thought as the major cause that most of the financial indexes tend to deteriorate at the same time when firm fails in addition to there exists strongly negative correlations among financial indexes.

5-3 Parameter Estimates

In this study, parameters are estimated by comparing all of the combination patterns of variables to decide the most suitable variable combination. However, this method does not always decide the optimum combination pattern of variables. Moreover, as the number of variables used in the model increases, the number of combination patterns of variables that have to be considered in the model increases rapidly. And as a result, considerable time is needed to find the optimum model parameters.

To respond to such issues, in the normal logit model or linear regression model, some statistical variable choice algorithms such as Stepwise/Backward/Forward are proposed. Regarding this point, the same type of variable choice algorithm is needed to find the most appropriate combination pattern of variables for the non-compensatory logit model. Further study is required to resolve these issues.

6. SUMMARY AND FUTURE INVESTIGATIONS

In this study, we proposed improving logit model based on non-compensatory rule instead of traditional normal logit model based on compensatory rule. We applied actual financial data for our model and showed superiority of our model.

Only financial ratios are used as the calibration data. But generally, when we apply default forecast model for small and medium entertainments or small individual firms, financial data may be window-dressed or the data may be modified intentionally. As a result, reliability for the financial data is reduced in such companies. To remove this issue, it is advocated that not only financial ratios but also individual attributes should be included in the model. However, including and combing such individual attributes is problematic. In an actual business, models are built by considering individual attributes and financial ratios simultaneously in the same logit model or mixing both of the estimated scores with a certain ratio after parameters are estimated independently on both of individual attribute data and financial ratio data. We can expect to build much stronger model through these processes.

REFERENCE

- [1] Beaver, William (1966), "Financial Ratios As Predictors of Failure", *Journal of Accounting Research*, 4(Supp),71-111
- [2] Altman, Edward, I., Financial Ratios (1968), "Discriminant Analysis and the Prediction of Corporate Bankruptcy", *Journal of Finance*, 23(4),589-609
- [3] Martin, Daniel (1977), "Early Warning of Bank Failure: A Logit Regression Approach", *Journal of Banking and Finance*, 1(3), 249-276
- [4] Dawes, Robin M and Bernard Corrigan (1974), "Linear Models in Decision Making", *Psychological Bulletin*, 81 (March), 95-106
- [5] Cattin, Phillipe and Dick R.Wittink (1982), "Commercial Use of Conjoint Analysis: A Survey", *Journal of Marketing*, 46 (Summer), 44-53.
- [6] Curry, David J., Jordan J. Louviere, and Michael Augustine (1981), "Comment on the Sensitivity of Brand Choice Simulators to Attribute Importance Weights", *Decision Sciences*, 12 (July), 502-516.
- [7] Einhorn, Hillel J., Donald N. Kleinmuntz, and Benjamin Kleinmuntz (1979), "Linear Regression and Process-Tracking Models of Judgment", *Psychological Review*, 86 (September), 464-485.
- [8] Einhorn, Hillel J. (1970), "The use of nonlinear, noncompensatory models in decision making", *Psychological Bulletin*, 73(3), 221-230
- [9] Goldberg, Paul E., J.Douglas Carroll, and Steven M.Goldberg (1981), "A General Approach to Product Design Optimization Via Conjoint Analysis", *Journal of Marketing*, 45 (Summer), 17-37
- [10] Wainer, H. (1976), "Estimating Coefficients in Linear Models: It Don't Make No Nevermind",

Psychological Bulletin, 83 (March), 213-217

AUTHORS BIOGRAPHIES

Yoshikazu, Sakamaki is a student of doctor course in Tokyo Institute of Technology, Department of Value and Decision Science, Japan. He received a B.Eng in the

Department of Control and System Engineering from Tokyo Institute of Technology and his M.Eng in the Department of Industrial Engineering and Management from Tokyo Institute of Technology. His recent work has involved financial engineering, marketing research with use of statistical approach. His email address is megacity@f3.dion.ne.jp .

Study of SOA Component Dynamic Scheduling Based on Mobile Agent Coalition

Tao He¹, Bin Tang², Leqiu Qian³

Computer Science and Engineering School, Fudan University, Shanghai 200433, China
¹eGrid@msn.com, ²edesk@126.com, ³iAgent@163.com

ABSTRACT

Service-oriented components differ greatly with the traditional ones in the Service-Oriented Architecture. The ways of scheduling components seamlessly according to the agile computing needs to fit the e-business requirements is the key technology in the highly distributed, paralleled environment. In this paper, Based on the Multi-Agent Coalition, a new service-oriented component dynamic scheduling model is proposed, including the Multi-Agent Organization to schedule and coordinate the component assembly, the design of virtual execution task list table and self-learning algorithm, the definition of the Services component model, and the mechanism of collaboration Agents to search, discovery, concurrent schedule, dynamic assembly when execution in an heterogeneous network environment. To a large extent, the thesis solves the traditional problem of over-emphasis on centralized control logic, which leads to lacking flexibility in e-Business computing presently, and helps e-business service-oriented components become more adaptive, mobility and intelligence.

Keywords: Service-Oriented Component, Multi-Agent, e-Business, Scheduling, Algorithm

1. INTRODUCTION

Automation is the most effective way to improve the efficiency of e-business. With the widely use of Service-Oriented Architecture software engineering, there is at least two problems emerged: too much service-oriented components which lack of automatic management system and too weak technology now-used which cannot support the dynamic and agile computing to meet e-business demand, especially to fully considerate the enterprise individuation. Multi-Agent technology as the research focusing on Artificial Intelligence and distribution computing, gives an effective way to solve these problems, which make business becomes more agility and interconnectivity.

A new service-oriented component-scheduling Model based on the Multi-Agent Coalition is proposed. In the architecture, every kinds of agent can enhance mobile agility by self-learning mechanism and by storage of personal information in every dynamic and distributed execution. This approach use a fully new method model to construct the virtual business activities by simulating transactions in the real business world, such as locate customs, negotiate with them, and so on. We use the model in the following by build a MAGE (Multi-Agent Environment) to put the conception into practice, and the process of e-business is simplified greatly.

2. DYNAMIC COMPOSITION

Agent-based component dynamic composition is the fundamental technique to shift from traditional components to virtual service-oriented ones. With the help of agents, components can aggregate or compose existing services and component into new styles. Some components are virtual ones when execution, in fact,

they don't exist at all. Such would allow third-parties to provide value-added components, which is similar with the real world services business. Five new features of services component are defined here.

Dynamic Existence It will not only being an entity as used to be, but a function correlation with the time.

Autonomy The Agent-based Services Component has the ability to take its own decisions in an independent way, including selecting the best of the available options according to e-Business computing requirements, Self-adaptation to different interfaces and protocols, being flexible, polytypic and extensible.

Competitiveness It should be competitive for a candidate to integrate into the SOA. More efficient and accuracy should be achieved with the lowest costs, and the self-learning mechanism which enhance the performance of components after execution is vital.

Sociability Communication protocol of different Agents is needed to enable them interaction with each other.

Semantic matching Semantic matching of component descriptions will make composition more transparent.

There are two basic composition styles under the control of composition Agent. One is Hierarchical Composition, and another is dialogic Composition.

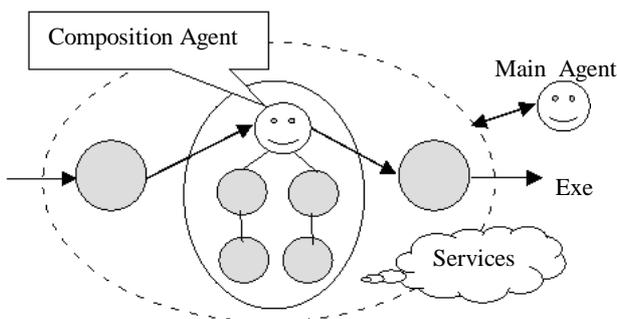


Fig. 1 Agent-based Hierarchical Composition

In hierarchical compositions, some related components cooperate under the agent coordination act as a new component at a higher level.

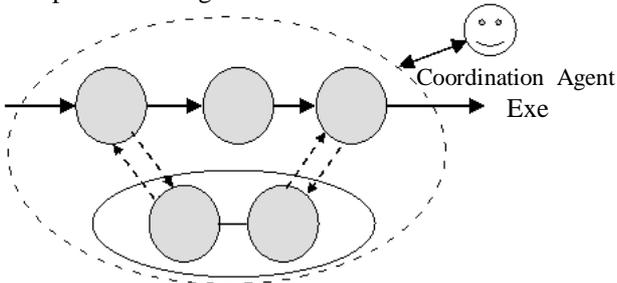


Fig. 2 Agent-based dialogic Composition

In the dialogic pattern, interacting components are viewed as peers. One component may make use of the other by exchanging data or control signals supervised by coordination Agent.

3. RUN-TIME ASSEMBLY MODEL

The service-oriented component dynamic scheduling and assembly model during the run-time should be predictable, applicable and manageable. We investigate the mechanism from three perspectives including Agent Coalition Perspective, Component Dynamic Assembly Perspective and the Inside Control Gene Perspective.

3.1 Agents Society Architecture in the Model

According to the dynamic assembly requirements, a main control-agent is introduced. It is an agent stored in migrating node with a virtual task list, and when the main agent runs in the migrating node, these service-oriented components can be assembled dynamically and the task table can be taken down after running at a previous node.

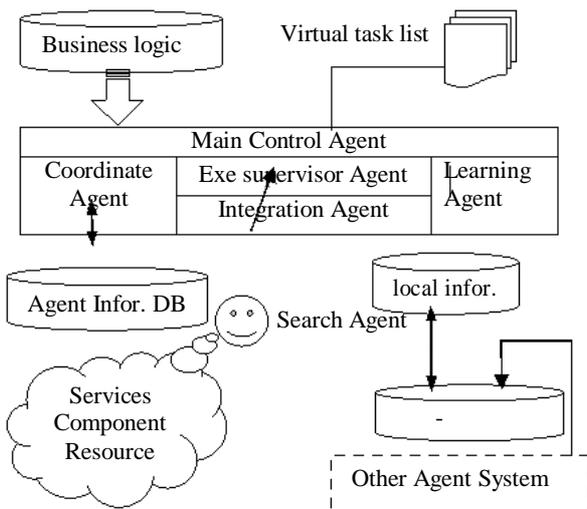


Fig. 3 Agents Society of Run-time Assembly Model

Secondly, the mobile Multi-Agent coalition is composed of some other kinds of agents including: services search agent, execution supervisor agent, learning agent,

integration agent, coordinate agent and safeguard agent, which make distributed services component composite, integrate, scheduling dynamically according to business logic. All of them function and cooperate with each other to overcome the shortcoming of traditional e-business system and create a real-time virtual cyber enterprise.

3.2 Component Dynamic Assembly Model

The Dynamic Assembly Model covers the whole process of the service-oriented components discovery, composition, negotiation, and orchestration dynamically and seamlessly, which makes it possible when fulfilling some complex commercial tasks by interacting with distributed component resources. Nowadays, many distributed service-oriented applications designed to execute on Grid Computing Environment, require the simultaneous co-allocation of multiple service-oriented component resources in order to meet high business performance frequently. The approach via agent coalition mechanism for services allocation and management architecture is very helpful to the co-allocation resource efficiently. With the help of Mobile Agents, new component clusters are generated and coupled to run at a higher level. So, the main focus is to design more adaptive component interfaces to enable the assembly and scheduling process more flexible and context-aware. Agent is the best solution.

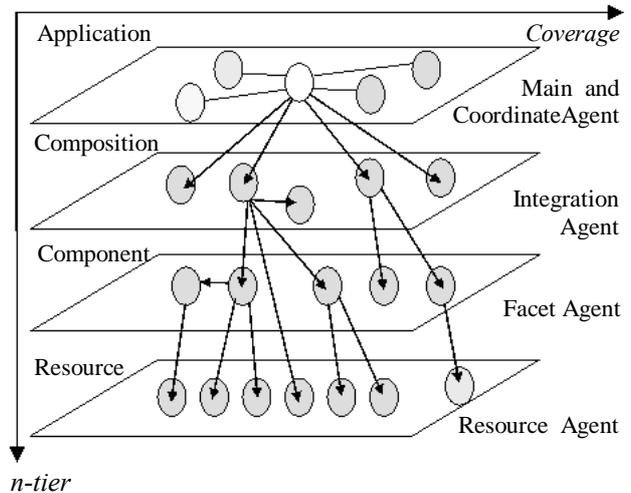


Fig. 4 Component Dynamic Assembly Model

The Facet Agent based on XML encapsulates the detail information of interface and was supported by the component assembly history information warehouse. It has self-descriptions and can communicate with other component agents to composite into a new style according to the e-Business requirements and the characteristic information retrieved from history data. Main Control Agent can migrate to other node to supervise the compiling and execution of local assembled component according to the Virtual Task List optimized by dynamic scheduling Algorithm. After each successful running, some key information about the component will be inserted to the knowledge base.

3.3 Run-time Agent Gene Design

A soft gene for agent is a hereditary unit that determines some particular characteristics in a component. After adopting certain soft genes, the software component can make certain practice activities, update its own behavior rules, etc. the main agent gene is the soul of the system in the run-time assembly process. It is define as XML, whose items can be easily updated by operation classes. We give the main definitions here:

```
<? xml Version="1.0" encoding = "utf-8">
  <Main Agent url=" " KnowledgeBase=" " ...>
    <Agent id=0001 type="ExeAgent">
      <Infor>It is a Exe Supervisor Agent!</Infor>
      <IP>12.121.10.156</IP>
      <KnowledgBase>url </KnowledgeBase>
      <DataPool>virtual execution lists</DataPool>
    </Agent>
    <Agent id=0002 type="SearchAgent">
      <Infor>It is a Search Agent!</Infor>
      <IP>12.121.10.100</IP>
      <KnowledgeBase>url </KnowledgeBase>
      <DataPool>search requirements</DataPool>
    </Agent>
    <Agent id=0001 type="CoordinateAgent">
      <Infor>It is a Coordinate Agent!</Infor>
      <IP>12.121.10.101</IP>
      <KnowledgeBase>url </KnowledgeBase>
      <DataPool>coordinate relations</DataPool>
    </Agent>
    <Agent id=0001 type="CompositeAgent">
      <Infor>It is a Composite Agent!</Infor>
      <IP>12.121.10.1</IP>
      <KnowledgeBase>url </KnowledgeBase>
      <DataPool>component id</DataPool>
    </Agent>
    .....
  </Main Agent >
```

Gene MainAgent

```
{ Attribute:
  double Skillness,Intelligence,LeaderAbility;
  string IPaddress, State, Task ;
  struct *VirtualExeList, *MessageQueue ;
  Agent *xAgent[N] ;
Method:
  InitiateAllAgent();
  UpdateGene(XML DOM)
  SendControlMessage(Agent *Target);
  CheckGoal(*Compnent);//whether meet the goal
  Scheduling(struct *VirtualExeList);
  MigrateToDestin(string IPadress);
  .....
  Destroy();
}
```

Gene ExeAgent

```
{ Attribute:
  long int TaskNum;
```

```
struct *VirtualExeList, *MessageQueue ;
.....
Method:
  OptimizeExeList(*VirtualExeList);
  UpdateGene(XML DOM)
  .....
}

Gene CompositeAgent
{ Attribute:
  long int ComponentNum;
  struct *Compo, *MessageQueue ;
  .....
Method:
  ComponentInsert(*Compo, TagentAddress);
  ComponenDelete(*Compo, SourceAddress);
  SearchRequest(string Requirement)
  UpdateGene(XML DOM)
  .....
}.....
```

With the help of Agent Gene revised and optimized adaptively, components can be easy understood, assembly according to the commercial computing requirements compared with that of before.

4. SCHEDULING ALGORITHM

4.1 Agent-based Concurrent Algorithm

Agent-based concurrent scheduling algorithm processes requests for service between agents and the executions of the tasks list. We mainly use the Enhanced Coffman-Graham Algorithm to perform concurrent scheduling to reduce the time of execution. Applicable components of systems are organized logically according to the optimized virtual execution task list.

:

Algorithm

Tasks List: $T = \{T_1, T_2, \dots, T_n\}$ $a(T_i): T_i \text{ Tag}$
 T_i successor aggregate : $S(T_i) = \{T_k | T_i \Theta T_k\}$
 $N_{EC}(T_i): \text{Max}\{\text{time}(T_i + S(T_i))\}$

Step1: Include tasks without successor and tag as $\{R\}$;

Step2: for $i=1$ to n do

- { a. Calculate $N_{EC}(T_i)$ of all T_j in $\{R\}$;
- b. take out the Task with $\text{Min}(N_{EC}(T_i))$;
- c. if $\text{Min}(N_{EC}(T_i)) > 1$, Random select a T_R ;
- d. $a(T_R) = i$;
- e. Add the task Which has no Tag but its successors are all tagged into $\{R\}$ in T_R preceding aggregate.

Step3: Construct $L' = (u_n, u_{n-1}, \dots, u_2, u_1)$, to ensure $a(u_i) = i$ ($i=1, 2, \dots, n$);

Step4: Eliminate virtual tasks in L' to construct L ;

Step5: To $(T \Theta L)$, use common Graham algorithm to scheduling tasks in L ;

Reference: common Graham algorithm^[7]

4.2 Mismatch Component Assembly algorithm

The semantic matching of service-oriented component is always hindered by mismatches when assembly. Formalized definition of matched component is:

a) $Match(O, O') = m_{spec}(O_{spec}, O'_{spec}) \wedge m_{sig}(O_{sig}, O'_{sig})$

If O and O' signature and specification match, then they are matched components.

b) R_j : the e-Business Required Components vector space

C : the exist component in service environment

$$M(C, R_j) = \frac{\sum_{i=1}^N r_{i,j} \times c}{\sqrt{\sum_{i=1}^N r_{i,j}^2} \times \sqrt{\sum_{j=1}^N c^2}}$$

If $M(C, R_j) < allow_{min}$ then they are match.

When the Search Agent find unmatched component, it can solve in the follow ways by coordinate agent:

- Send a message to Main Agent to adjust to suit the component if the execution can meet the goal
- Search another component to composition as a matched one
- Use Interface Agent to mask unnecessary function and parameters
- Function as middle ware to cover the discrepancy
- Duplicate and revise the original code to produce a new suitable component according to knowledge base.
- Specialized the component some control parameters to get the sub-function of a component
- Use learning mechanism to enhance the component function to match the requirement
- Translate the component I/O data format

4.3 Run-time Agent Self-learning Algorithm

Collaborative and Shareable Learning System is used to improve the performance of Mobile Agent Coalition on service-oriented components dynamic scheduling. It's the integration of Reinforcement Learning, Bidding, Genetic and other Algorithms.

We use shared knowledge and commercial logical by refining each agent's particular knowledge to improve the assembly of components and scheduling of task. During the run-time, the agents can dynamically alter the learning contents and presentation over time. Specifically, the agents can migrate and participate in a number of important tasks such as automatic composition of component according to the commercial logic, analysis and delivery services to users and adaptation to new patterns and services. This self-learning algorithm system is hierarchical in terms of its distribution to different agents. Agent Coalition can use learning mechanism to upgrade other algorithms includes component evolution and aberrance algorithm, virtual task list optimization algorithm, and commercial logical pattern auto-builder algorithm.

Every type of Agents can have its own and common knowledge base, locally or remotely, to build a

cooperative, co-evolutionary mechanism in which agents can learn by using a team-based reward system. System will obtain the best team for achieving the task in e-business computing environment which requires coordination to succeed. These agents mainly use the auction mechanism to negotiate their roles dynamically. Agents bid individually according to their perceptions. The system chooses the best combination of bids for the team; the chosen bids may not be optimal for each individual, but the system learns to bid as a team and develops the best team-based strategy. We develop our learning algorithm in a distribute, simulated version of the real environment intelligence

5. CONCLUSION AND FUTURE WORK

Services-oriented Component based on Multi-Agent is the revolution to the traditional computing technologies, which make the information infrastructure be more quickly adapt to changing business priorities. Here we propose the framework and main part algorithms' brief view, which is an effective way to achieve our goals. But, some algorithms is far more satisfied, and still need to give deeply research for better solutions, especially for more matured Self-learning, effective assembly and efficient scheduling Algorithms.

ACKNOWLEDGEMENT

This work has been supported by the National Natural and Science Foundation of China granted in this year.

REFERENCES

- [1] S.McIlraith, T.C.Son, "Semantic Web Services". *IEEE Intelligent Systems, Special Issue on the Semantic Web*, 16(2): pp46-53, March/April 2001
- [2] Laurent, "Component-oriented software technology". Niestrasz, "Object-oriented Software Composition", Prentice Hall International, pp3-28, 1995
- [3] R.H.Reussner. "The use of parameterized contracts for architecting systems with software components". In W.Weck, J.Bosch, and C.Szyperski, editors, *Proceedings of the Sixth International Workshop on Component-Oriented Programming*, June 2001
- [4] "Business Process Modeling Language", Accessed June 2004 from <http://www.bpmi.org>
- [5] Van der Aalst WMP. "Petri Net Based Workflow Management Software". *Proceedings of the NFS Workshop on Workflow and Process Automation in information System*. Athens, Georgia, May 1996.
- [6] Lawrence Wilkes, "Web Services Roadmap for On Demand Business", *CBDJ Journal*.
- [7] Daniel, Jon Feldman, "Parallel Processor Scheduling with Delay Constraints", in *Proc. Of the SAINT Symposium*, Jan.2001
- [8] Luck, "Multi-Agent Roadmap". www.AgentLink.org

Temporal Association Rule Mining in China's Closed-end Fund Data

Ying Liu, Guoqing Chen

School of Economics and Management, Tsinghua University, Beijing 100084, China
{liuy6, chengq}@em.tsinghua.edu.cn

ABSTRACT

Financial market plays an important role in economy. Although funds developed only a few years in China, it has been a focal point in research and practice. The conventional methods analyzing fund data are fundamental analysis and technical analysis. Data mining can extract implicit, previously unknown and potentially useful knowledge from data. This paper presents the new technique to analyze China's closed-end fund data and temporal association rules (TAR) are discovered which reflect the relationship among open price, close price, trading volume and grail index. Experimental results show some interesting outcomes.

Keywords: data mining, temporal association rules, closed-end fund

1. INTRODUCTION

Fund is an important investment tool. During the past several years, fund market has developed rapidly in China. By the end of 2003, there were totally 50 fund management companies issuing 54 closed-end funds and 56 opened-end funds. Closed-end funds had reached 86.2 billion RMB by net assets and opened-end funds had been up to 76 billion RMB by circulation [1]. There are at least two types of methods, namely fundamental analysis and technical analysis, for analyzing fund data, which are widely recognized and applied. Fundamental analysis typically focuses on the general economy, the condition of the industry and the condition of company. Technical analysis is an approach to stock investing where the past prices are studied and using charts as the primary tool [2].

Data mining is to discover interesting patterns from large amounts of data. Data mining functionalities include the discovery of association, classification, prediction, clustering and so on. As a promising technology, which may be supplementary to conventional methods, data mining in financial data has increasingly attracted attention of academia and practitioners. This paper focuses on finding potential association from closed-end fund data which is deemed helpful in fund investment decision-making.

2. TEMPORAL ASSOCIATION RULE

Association rule is one of the most important forms of knowledge [3]. It is of the form $X \Rightarrow Y$, where X and Y are item sets in databases. Two thresholds, namely degree of support and degree of confidence, are defined to evaluate the rule. For example, rule: trading volume rises \Rightarrow close price rises, reflects a sort of association relationship between price fluctuation and trading volume. Apriori algorithm is regarded a classical algorithm [4] for association rules mining. Several efforts have been made to extend the classical association rule mining to inter-transactional association rule mining [5] and certain applications have been proposed in stock movement predictions [6].

Notably, fund data is typically time-series, which results in the need of temporal association rules mining. A Temporal Association Rule (TAR) is of the form

$X \xrightarrow{t} Y (t \geq 0)$ where X and Y are item sets, and t is the time-lag [7]. TAR is an extension of a conventional association rule, and represents that Y will occur at t time units after X occurs. Its degree of support (Dsupp) and degree of confidence (Dconf) are:

$$D \text{ sup p } (X \xrightarrow{t} Y) = \frac{\| X \cup Y \|}{|D| - t}$$

$$D \text{ co n f } (X \xrightarrow{t} Y) = \frac{\| X \cup Y \|}{\| X \|}$$

where $\|X\|$ is the number of records containing X at time point T, $\|X \cup Y\|$ is the number of records that contain Y at time point T+t in correspondence to the records containing X at T, and $|D|$ is the number of records in dataset D.

Chen et al [8] proposed an algorithm so as to obtain a balance between the semantics and complexity. Two types (i.e. upward and downward) of the stock returns patterns are considered, which result in Boolean data items. Furthermore, in many real world applications, discovering associations for multiple stock returns patterns (e.g., categorical with 3 values: upward, unchanged, downward) is meaningful in supporting financial decision-making. Moreover, for continuous returns (e.g., return values in continuous domains), the respective domains can usually be discretized into several intervals. This method is further investigated and applied to closed-end fund data in China in this paper.

3. TAR MINING IN CLOSED-END FUND DATA

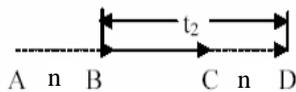
The characteristic of fund data is numerical, enormous and time-series. Our data for 2000-2003 come from the security database of TianXiang investment consulting. Every fund has its individual IPO time which results in different data volume. In order to assure the same external environment (i.e., policy, macroeconomics), selecting complete data in a same interval is necessary. Data from 2000.1.4 to 2002.12.31 were selected as the mining set and data from 2003.1.1 to 2003.12.31 will be used for test purposes. There are totally 20 funds that have complete data during this period in China market.

Let $p(T)$ and $p(T-n)$ denote the values of attribute p (i.e., p can be open price, close price, trading volume or grail index) of some fund on time points T and $T-n$ (e.g., unit of T is day). The n -day rate of increase $P_n(T)$ is defined as in (1).

$$P_n(T) = \frac{p(T) - p(T - n)}{p(T - n)} \quad (1)$$

Theoretically the domain of $P_n(T)$ is $[-1, +\infty)$. For association rule mining purposes, numerical data need to be transformed into discrete data. Thereby what we want to discover is the temporal association rules between rates of increase for funds. Note that n is the time units during which the rate changes for a fund, and t is the time-lag for association rules. Figure 1 is an illustrative description for n and t .

Figure 1: Description of n and t



Let $\max(p)$ and $\min(p)$ respectively denote the maximum and minimum value of attribute p . K denote the number of discretized intervals. Then c denotes the interval span which is defined as in (2).

$$c = \frac{\max(p) - \min(p)}{K} \quad (2)$$

Let $q_1, q_2, \dots, q_{(K-1)}$ represent $(K-1)$ points used to discretize the domain of $P_n(T)$. Then the values of $P_n(T)$ can be discretized into K intervals, where q_i ($i=0, 1, 2, \dots, K$) is defined in (3).

$$q_i = \begin{cases} -1 & i=0 \\ \min(p) + i \cdot c & i=1, 2, \dots, K-1 \\ \max(p) & i=K \end{cases} \quad (3)$$

When $P_n(T)$ belongs to $[q_i, q_{i+1})$, it is labeled as $(i+1)$.

An example of the process is illustrated in Table 1 as follows with $K=4, n=1, t=1$. The values in Table 1 represent data of fund 500001 from Jan, 1st, 2000 to Jan, 10th, 2004 including open price, close price, trading volume and grail index.

Table 1: Data of fund 500001

date	open price(¥)	close price(¥)	trading volume(share)	grail index
2000-01-04	1.17	1.19	250,737	1,494.50
2000-01-05	1.2	1.17	228,309	1,498.05
2000-01-06	1.17	1.21	210,202	1,555.95
2000-01-07	1.21	1.25	628,288	1,611.81
2000-01-10	1.26	1.25	455,134	1,642.53

Each value in Table 2 represents the corresponding 1-day rate of increase of open price, close price, trading volume and grail index calculating by (1).

Table 2: Rate of increase of fund 500001

date	Rate of increase of open price	Rate of increase of close price	Rate of increase of trading volume	Rate of increase of grail index
2000-01-05	2.564103	-1.68067	-8.94483	0.237538
2000-01-06	-2.5	3.418803	-7.93092	3.865025
2000-01-07	3.418803	3.305785	198.8973	3.59009
2000-01-10	4.132231	0	-27.5597	1.905932

The discretization values of the rates are represented by integers in Table 3.

Table 3: Discretization data of fund 500001

date	Discretization of open price	Discretization of close price	Discretization of trading volume	Discretization of grail index
2000-01-05	3	1	1	1
2000-01-06	1	4	1	4
2000-01-07	4	4	4	4
2000-01-10	4	2	1	3

The price of funds fluctuates greatly at ex-dividend data. It is usually regarded as a kind of noise which needs to be filtered out.

In Table 4, each of the first five columns represents fund 500001's discretization value of every attribute. The last two columns represent temporal data of open price and close price. Considering these temporal data as new attributes, temporal association rules can be obtained by Apriori algorithm on the new table. For example, if in this table, rule: *grail rises* \Rightarrow *temporal open price rises* can be discovered, then it is a rule with a time-lag of 1 day. A detailed description of the results can be seen in section 4.

Table 4: Fund 500001's discretization data added temporal attributes

ID	Open price	Close price	volume	grail	Temporal open price	Temporal close price
1	3	3	1	4	1	4
2	1	4	1	4	3	4
3	3	4	4	3	3	3
4	3	3	3	1	2	2
5	2	2	2	2		

4. RESULTS

Temporal association rule mining technique has been applied to closed-end fund data in China and a few interesting temporal association rules have discovered which reflect the potential knowledge hidden in the fund data.

Concretely, 10 rules with high level of D_{supp} and D_{conf} are discovered. Most of the rules are consistent with those with conventional methods. However, some of the rules discovered are new that cannot be easily obtained by conventional methods including fundamental

analysis and technical analysis.

4.1 Relationship among funds

According to the results of experiments, all funds support the rule $grail\ rises \Rightarrow open\ price\ rises$ with both high Dsupps and Dconfs, except for the funds with small sizes as 1 to 1.5 billion RMB. In addition funds issued by the same fund management company have similar Dconf with grail index.

First, we display the association between fund price and grail index. Over half of the rules are involved in 80 percent of the twenty funds, and parts of them are consistent with those with conventional methods. Rules are listed in Table 5. The result indicates that funds are very similar with each other. It is supposedly relevant to great similarity in stock-pick, means of manipulation, stock centralized degree, stock position noted in quarterly reports of each funds. All of these may lead to similar rules among funds and the extremely trend-similarity between funds with grail index.

Especially, it also can be noticed that small funds with 1 to 1.5 billion RMB did not follow the same pattern $grail\ rises \Rightarrow open\ price\ rises$, which indicates that small funds operating with the limited capital have to keep centralized stock investment, and thus behave less dependent on grail index.

Table 5: Temporal association rules

Temporal association rules: (minsupp=0.8 minconf=0.8)	
$grail\ rises \Rightarrow open\ price\ rises$	
$open\ price\ rises \Rightarrow grail\ rises$	
$volume\ falls \Rightarrow close\ price\ falls$	
$grail\ falls \Rightarrow close\ price\ falls$	
$volume\ falls, grail\ rises \Rightarrow open\ price\ rises$	
$grail\ rises, open\ price\ rises \Rightarrow open\ price\ rises$	
$close\ price\ rises, grail\ rises \Rightarrow open\ price\ rises$	
$volume\ falls, grail\ falls \Rightarrow close\ price\ falls$	
$grail\ rises, close\ price\ rises \Rightarrow volume\ rises$	
$volume\ falls, open\ price\ falls \Rightarrow close\ price\ falls$	

Thus, the investment strategy of the small-sized funds is different from that of larger funds. Small funds may be more independent and their stock-picks are of more concern. When it comes to large funds, the focus should be on the present price and the volume of grail index should be focused rather than on a certain stock, since there exist strong association between larger funds and grail index.

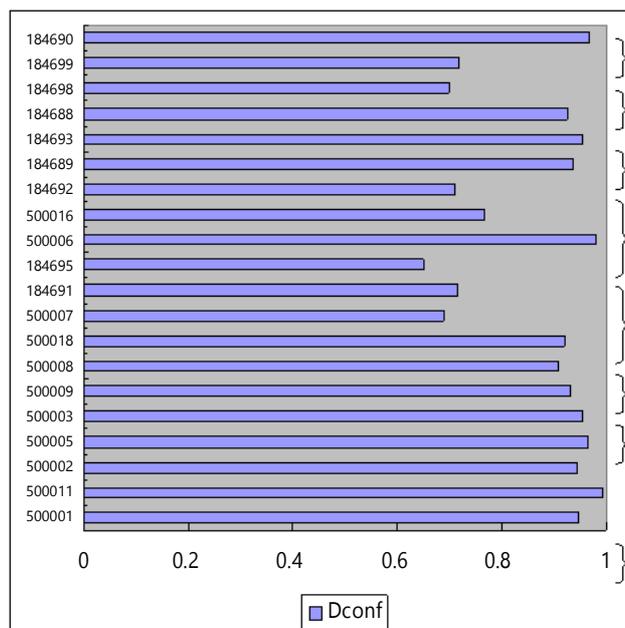


Figure 2: Dconfs of rule: $grail\ rises \Rightarrow open\ price\ rises$

Second, Dconfs among 20 funds are compared. Figure 2 shows the Dconfs of rule: $grail\ rises \Rightarrow open\ price\ rises$ of all funds grouped by the fund management company. Funds which pertain to the same fund management company are indicated by parenthesis. It is obvious that most funds under the fund management company have similar values of Dconf. It implies that each fund management company carries out similar operation modes and manipulation ideas on its funds. The Dconfs of the funds in a certain fund management company are similar in most situations despite of different fund managers, which show that the characteristic of the manager is not as remarkable as that of the company. Figure 2 also reveals that it did not seem helpful to reduce the risk by purchasing different funds of the same company because of the similarity of the funds.

4.2 Price trend of funds

If both minsupp and minconf are set to 0.9, the rule: $close\ price\ rises \Rightarrow close\ price\ rises$ can be found in 8 funds which are listed in Table 6. The rule indicates that the higher the Dconf is, the longer that price keeps rising. Notably, six of the eight funds are also ranked very high based on the rate of increase of net asset.

Table 6 Rule: $close\ price\ rises \Rightarrow close\ price\ rises$

Fund code	Dsupp	Dconf	Rank by rate of increase of net asset	Rate of increase of net asset
184690	0.942	0.956	1	29.99%
500003	0.915	0.932	2	29.93%
500002	0.904	0.921	3	28.66%
500011	0.972	0.977	5	21.44%

184693	0.929	0.939	7	19.17%
500006	0.957	0.967	8	17.82%
184689	0.905	0.913	12	14.63%
500001	0.915	0.935	13	14.45%

Results also show that more than 90% funds remain the same in content of association rules as time-lag changes, although Dconf changes in a small range. Experiments were carried out by setting time-lag as 1, 2, 3 days respectively.

Take Fund TongYi (184690) as an example. The results are listed in Table 7.

Table 7: Results of different temporal time

rule	t=1	t=2	t=3
grail rises \Rightarrow open price rises	0.925	0.900	0.899
Close price rises, grail rises \Rightarrow open price rises	0.939	0.910	0.898
volume falls, grail rises \Rightarrow open price rises	0.928	0.899	0.896
grail rises, open price rises \Rightarrow open price rise	0.935	0.908	0.909
grail rises, close price rises \Rightarrow volume rises	0.956	0.953	0.9595
volume falls, grail falls \Rightarrow close price falls	0.889	0.887	0.886
grail falls \Rightarrow close price falls	0.891	0.892	0.891
volume falls, grail falls \Rightarrow close price falls	0.893	0.889	0.890
volume falls, open price falls \Rightarrow close price falls	0.895	0.895	0.899

Although values of Dconf change a little, they are still high, which implies that price trend is stable.

Investment suggestions may be summarized from above analysis: synthetically considering macroeconomic situation and trend of grail index, buying in and holding funds once the price of one of the eight funds rises, because the price of these funds will continue rising.

4.3 On test data

Data from Jan, 1st, 2003 to Dec, 31st, 2003 are used for test purposes. Because of space limitation, only two of the rules supported most in 20 funds are tested and listed in Table 8.

Table 8: Test results

Fund code	Dconfs of rule1	Dconfs of rule2
184688	0.85	0.86
184689	0.8	0.81
184690	0.71	0.74
184691	0.89	0.89
184693	0.87	0.88
184695	0.83	0.87
184698	0.83	0.83
184699	0.82	0.81
500001	0.79	0.8
500002	0.8	0.79
500005	0.81	0.82
500006	0.89	0.89

500007	0.79	0.8
500008	0.95	0.94
500009	0.8	0.73
500011	0.9	0.9
500016	0.71	0.73
500018	0.95	0.95
Average	0.83	0.83

These two rules are *grail rises* \Rightarrow *open price rises* and *volume falls, grail rises* \Rightarrow *open price rises*. The values of Dconfs are both averaged above 0.80. It demonstrates that funds followed the same patterns in long term, as the rules derived from the first three years data were still supported by the fourth year data.

5. CONCLUSION

Nowadays, fund market plays a more and more important role in China economy. This paper introduced a new technique, temporal association rule mining, into the research of predicting fund return. Different from conventional methods, temporal association rule mining provides a data-driven method to discover interesting patterns in large-scaled fund data, which are previously unknown and potentially useful. The experiments on real closed-end fund data have shown that the method of mining temporal association rules in fund data was effective and could be used to support investment decisions.

ACKNOWLEDGEMENT

The work was partly supported by the National Natural Science Foundation of China (70231010, 70321001)

REFERENCES

- [1] <http://finance.sina.com.cn/yched/20040105/1054589797.shtml>
- [2] R. Edwards and J. Magee. Technical analysis of stock trends. *American Management Association*, 1997
- [3] Jiawei Han, Micheline Kamber. Data Mining Concepts and Techniques. *Higher Education Press*, 2000
- [4] Rakesh Agrawal, Tomasz Imielinski, Arun Swami, Mining Association Rules between Sets of Items in Large Databases, *Proceedings of the 1993 ACM SIGMOD Conference* Washington DC, USA, May 1993
- [5] Tung, K.H., Lu, H., Han, J., Feng, L., Breaking the Barrier of Transactions: Mining Inter-Transaction Association Rules, *the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD99)*, San Diego, CA, August 15-18, 1999
- [6] Lu, H., Han, J., Feng, L., Stock Price Movement Prediction and N-Dimensional Inter-Transaction Association Rules, *1998 ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD98)*, Seattle, Washington, June 5, 1998
- [7] G. Das, King-Ip Lin, H. Mannila. Rule Discovery from Time Series. *KDD98*, 1998
- [8] Chen G Q, Ai J, Yu W., Discovering Temporal Association Rules For Time-Lag Data. *In Proceedings of International Conference on e-Business (ICEB2002)*, 2002. 324

The Application of RFID in Emergency Medicine

Tsai W.C.¹, Frank Kuo², Yuehching Lee², F. Wu²

¹Institute of Health Service Management,
Taichung, China Medical University, Taiwan, 404, China

²Institute of Healthcare Information Management,
National Chung Cheng University, Chia-Yi, Taiwan, 621, China
wtsai@mail.cmu.edu.tw, {frankk, xkevin, kfwu} @mis.ccu.edu.tw

ABSTRACT

Recently natural disasters and man-made calamities happened very frequently in Taiwan, some of which caused hundreds of inhabitants died and injured. The most important thing that makes the emergency medical treatment system work effectively during these catastrophes is dependent on the accuracy and immediacy of the information among ambulances, hospitals, fire bureaus, public health bureaus or offices, and emergency executive centers. With this information, the medical system can arrange the resources efficiently.

Radio frequency identification (RFID) is a kind of wireless technology that it uses radio frequency (RF) to identify target objects. The characteristics of RFID are its identification and orientation. Through tracing electronic signals, RFID is able to identify, orientate and trace patients or medical materials directly and continuously. Thus, RFID can help managers of medical institutions to promote emergency medical treatment quality, reduce medical mistakes effectively, and increase efficiency and effectiveness in hospitals.

A series of emergency medical delivery applications that use 915MHz RFID rings is developed in this paper. Through the help of RFID, the paper integrates the information of the whole emergency medical supply chain, and establishes a new model of emergency medical activities to solve the problems of information blocking among different medical rescue units. By solving these problems, the new model can enhance the rescue quality, reduce the damage of the disaster and strengthen the patient-oriented emergency care system.

Keywords: Emergency medicine, RFID (Radio Frequency Identification), data exchange, patient reference. emergency care quality

1. INTRODUCTION

Natural disasters and man-made calamities happened very frequently in Taiwan in recent years. From the 921 earthquake in the middle of Taiwan and the typhoons from HERB, TORAJI to MINDULLE, these disasters had caused a lot of serious damages. After the accidents, we found several faults through the investigation of the emergency medical rescue system when facing these catastrophes. For example, the uninformed distribution of the patients, the lack of the concordance among the departments of the government, inappropriate allocation of resources, the unknown injured degrees of the patients, and the un-reach of the right and instant information to the officers of disaster rescue teams who have been recognized as major faults [3-6].

It is needed for the government to establish a more efficient system to handle such a situation when disasters happen. The system is useful in the processes of the crisis management, especially when the communication of messages is crucial and the integration of the related medical resource is needed. If the advantage of the information technology is suitably applied, it is expected that the rescue team can provide the health service in the first time and then reduce the damage of the major disasters.

RFID is known as the one of the top ten important technologies in this century [7], since of its identification and orientation characteristics. RFID, if applied in medicine, can provide the tracing of patients or medical materials directly and continuously. Thus, RFID can help the managers of medical institutions to enhance its emergency medical treatment quality, reduce medical mistakes, and increase efficiency and effectiveness in hospitals [1].

When crucial disasters take place, how to deliver information quickly and accurately among ambulances, hospitals, fire bureaus, public health bureaus, and emergency executive centers is a major key to save the lives. This paper is to design and apply RFID to the emergency medicine. The application of RFID in emergency medicine can be the process from the time when the patient in the ambulance to the time when the patient is moved to the emergency room. Through the whole process involved, the application of RFID can integrate the logistics of the emergency medicine and provide a solution for the questions that they can not be solved in the time that the information count not be shared and integrated.

The purpose of the paper is to aid rescue systems to

obtain the newest information for the different level of stages, namely, the operation stage (such as ambulances and hospitals), the management stage (such as fire bureaus and public health bureaus) and the strategic stage (such as control centers). In the operation stage, RFID system can utilize the medical resources according to the deployment of these resources and the injured degree of the patients that are classified before arrive the emergency room. In the management stage, RFID system can improve the coordination of data exchange and reduce the problems of disagreed information among the hospitals in responsibility, fire bureaus and public health bureaus. In the strategic stage, RFID system can help the directors of the government to get the condition of the casualties in disaster event, and help the directors to make adequate decisions to rescue victims of the disaster.

2. RFID EMERGENCY MEDICAL SYSTEM DEVELOPMENT

2.1. System development

The proposed RFID emergency medical system develops and designs based on the related technique specified in the white paper in EPC Network [8-10]. We designed the system with the method of prototyping. The users' satisfaction and demands were considered in the developing process. The emergency rescue process and its flow chart are showed in Figure 1.

The rescue process contains six steps, as is discussed in the following:

1. The people who witness or encounter the disaster event call or inform the local fire bureau whatever they have seen;
2. The commander of the fire bureau dispatches fire-fighting trucks immediately and notifies the related hospitals to send ambulances to the spots to offer emergency medical treatment;
3. The ambulances arrive at the spots to salvage the victims and pick them up to these ambulances;

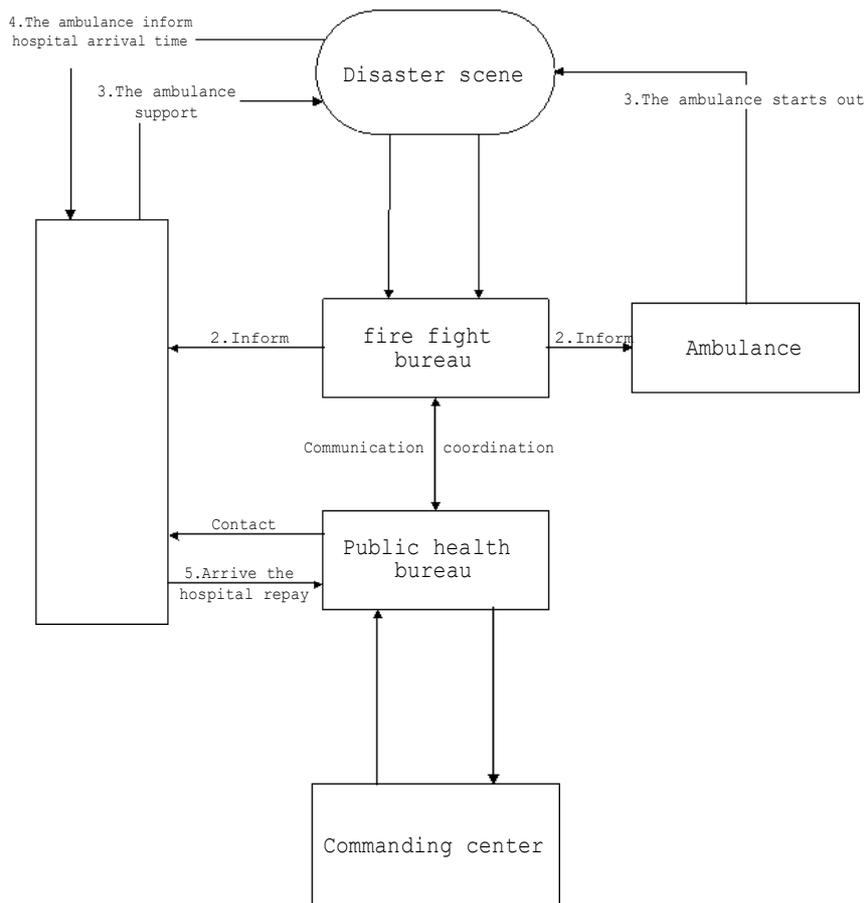


Figure 1. The flow of emergency medical rescue.

4. Before arriving the hospital, the rescue team notifies the hospital in the ambulances about their arrival time and the preparation of the treatment;

5. When the ambulance arrives to the hospital emergency room, the hospital will notify the public

health bureau of the patient's information. If the public health bureau doesn't receive the information, it can inquire the hospital. The information between the public health bureau and the hospital must be coordinated with each other;

6. The commander in the public health bureau notifies the command center about the information of the disaster. The commander in the command center arranges the relative units to deploy the resources to the injured patients.

According to the analyses of emergency rescue process,

we designed the emergency medical treatment with the introduction of RFID. The modified rescue system flow represented in a sequence diagram is showed in Figure 2. Then we program the framework for the RFID emergency medical rescue system, which is divided into the operation, management and strategic stages, as is shown in Figure 3.

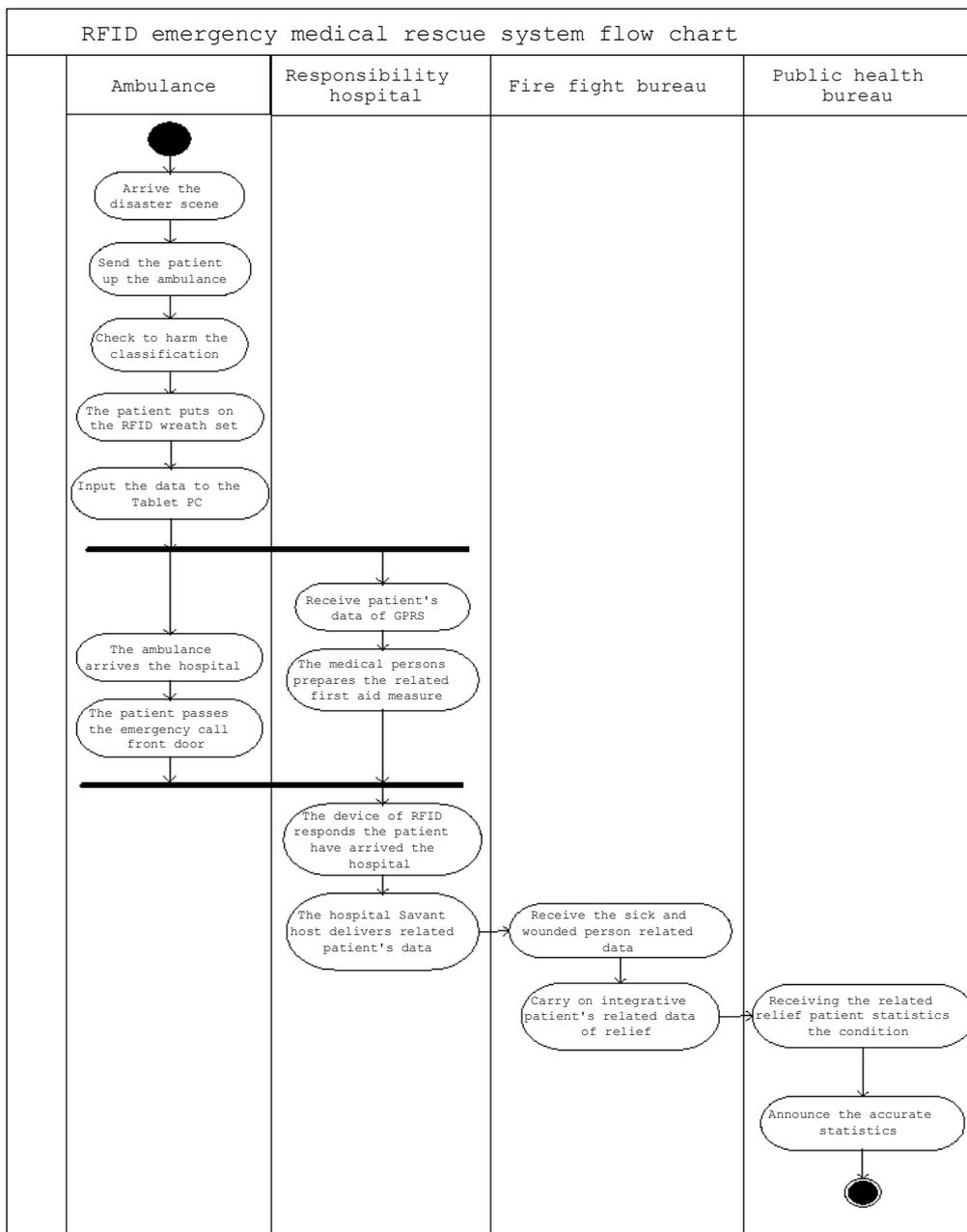
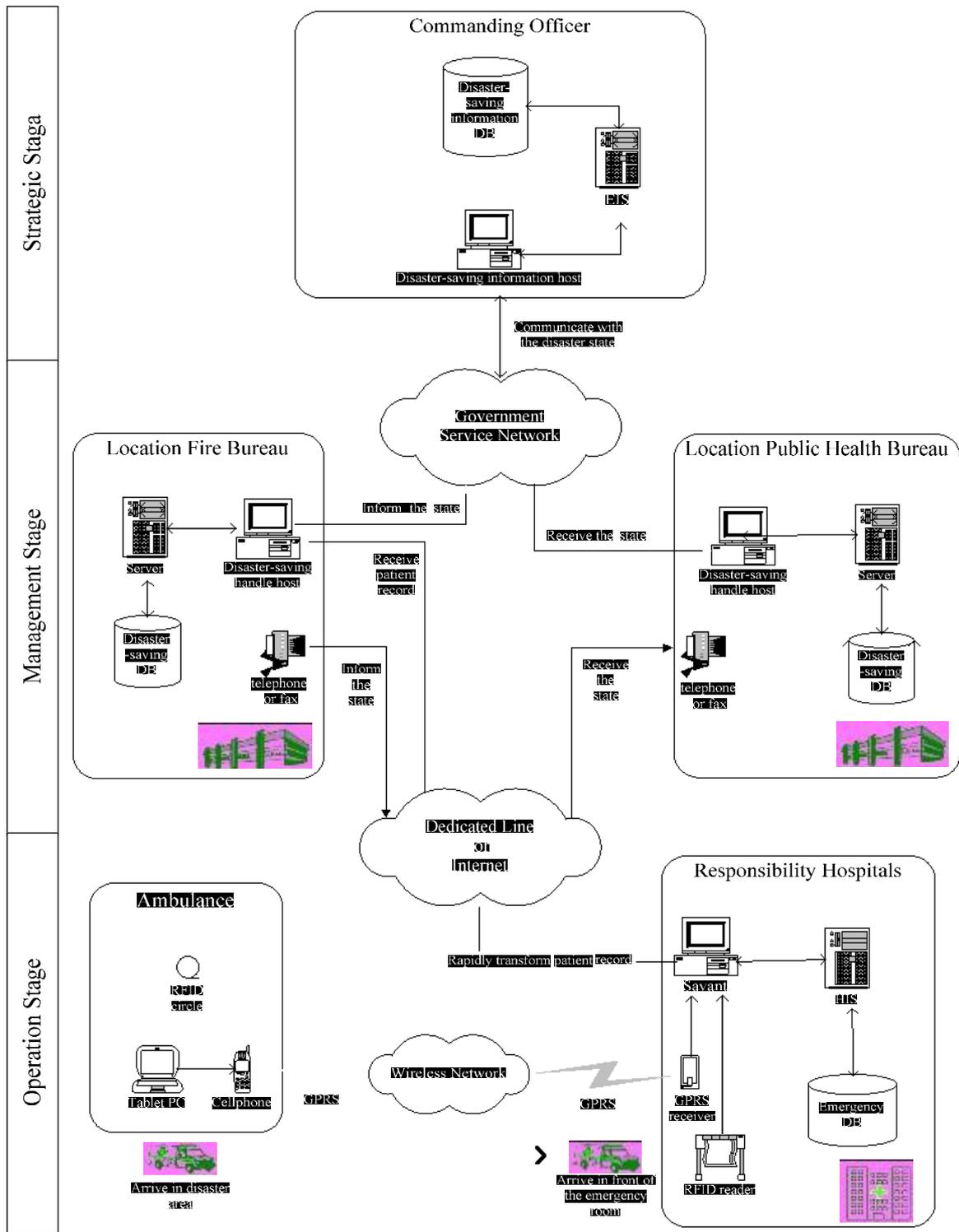


Figure 2: RFID emergency medical rescue system flow chart



middleware, which handles the RFID events, carrying out the extraction and summary of data and translating the information such that the information can be processed by other computer systems;

3. The middleware also integrates the Savant server and the HIS's database. The middleware transfers data in XML format to call other system components that is encapsulated in the Web Service. Using SOAP standard, the system integration is done.

3. BENEFIT ANALYSIS

The development of RFID emergency medical rescue system follows the guideline to establish an information system. The process of system analysis, system design, the hardware installation and the system construction can be easily achieved. In addition, we need to evaluate the system from four dimensions, namely, practicability, accuracy, efficiency and completeness:

1. Practicability

The feasibility, affordability and the future extension of the proposed system can be validated by the three facts: (1) the wider and wider coverage as well as the good quality of the mobile telecommunication and wireless Internet access system, (2) the appearance of the acceptable handwriting input system such as tablet PC and (3) the maturity of EPC Network, system have been developed maturely

2. Accuracy

The RFID tags are independent with each other. All rescue units can receive the messages sent from the tags accurately and automatically. Compared to human-entry system, the automatic operation system can reduce the risk of mistakes and malpractice;

3. Efficiency

The efficiency of the RFID emergency medical rescue system can be discussed in three viewpoints:

(1) The viewpoint of healthcare providers:

The system can directly connect to the HIS of hospitals. The system provides an easier and faster way to key in and recognize the patient in arranging the process of medical treatment. With this system, the healthcare providers can alleviate the stuffs and focus on saving the lives;

(2). the viewpoint of fire bureaus and public health bureaus:

Through the system, the local fire bureaus and public health bureaus can correctly and immediately allocate the available medical resources to the patients. In addition, the communications between them will be

smoother than ever;

(3). the viewpoint of the governments:

By the precise information, the leaders of the government can promptly make decisions to rescue lives. Even when the local salvage resources are run out in a big casualty, the leaders can be aware the condition immediately and then ask the aids from adjacent areas or countries to provide the sufficient medical treatments;

4. Completeness

As a result of implementing the RFID emergency medical rescue system, the county government can get over integrated information about emergency healthcare demand quantity and the number of casualties whenever catastrophes happen.

4. CONCLUSION

We have designed and implemented the RFID emergency medical system for the emergency medical rescue system. The motivation is to design a complete process from the time when the patients are taken into ambulances to the time when they arrive to the emergency rooms. The primary characteristics of RFID can provide the function of identification, orientation and traceability of patients. The communication of information at anytime and anywhere will make the emergency medical rescue more efficient. The proposed system offers a chance to integrate the whole medical supply chains and solve the medical information flow problems that can't be effectively shared and integrated before.

Through the development of the RFID emergency medical rescue system, we establish a better model of emergency medical rescue system. With the successful experience of a pilot experiment in Nantou, Taiwan, we established the relative performance indexes, models and standard operating procedures, and will transfer the successful paradigm to the other seventeen local emergency medical rescue networks. Through the integration of the object (i.e., patient) flows traced by the RFID and the information flows exchanged in HL7 standard, we can connect each local medical network and establish an emergency medical information integration network. The system can promote rescue quality, decrease medical mistakes, diminish the damage of disasters and finally build a patient-oriented healthcare environment.

5. ACKNOWLEDGEMENT

Thanks for the direct and inspire from the professor Hsin-Ginn Hwang, Cheng-Yuan Ku, and I-Chiu Chang in National Chung Cheng University (NCCU).

And special thanks for the professor Yi-Long Huang, Chuan Yi Tang, James Tsemeng Lin in National Tsing Hua University (NTHU); the professor Anthony Fu-Wah Han, Hsiao-Cheng D. Yu in National Chiao Tung University (NCTU).

Thanks the support from dear friends in 349-1 Lab in NCCU, and in "Green On Inc."

REFERENCE

- [1] Frank Kuo, Yueh-Ching Lee, Shi-Kun Ke, Chih-Ming Lee, Fan Wu, "Research issues in health industry with RFID," in Proceeding of the 3rd Asia-Pacific and Cross-Strait HL7 Conference on Healthcare Information Records, pp.148~160. Taipei, Taiwan, July 9~10, 2004.
- [2] Chen Yung Fu etc, The disaster in the Taiwanese region meets an emergency to handle the project and actual situation difficulty, The Department of Health science and technology project.
- [3] Wang Li Min, Hu Sheng Chuan, Disaster medical science, CHIN MING Publisher, 1999.
- [4] Chou Chih Chung, The Taiwanese urgent medical treatment gives first-aid the ten years to look back, The decade of the national urgent medical treatment business looks back the seminar thesis selected works, Taipei, Department Of Health, Taiwan, 2001.
- [5] Ma Hui Ming, The emergency medical system in the 21 st century, Journal of the Formosan Medical Association 4;215-17, 2000.
- [6] D.R. Reynolds, J.R. Riley, Remote-sensing, telemetric and computer-based technologies for investigating insect movement: a survey of existing and potential techniques, Computers and Electronics in Agriculture, 35; 271-307, 2002.
- [7] Klaus Finkenzeller; translated by Richel Waddington, "RFID Handbook- Radio Frequency Identification Fundamentals and Application", John Wiley, 1999
- [8] ID TechEx, <http://www.idtechex.com>
- [9] RFID Journal, <http://www.rfidjournal.com/>

The Design of a Web Document Snapshots Delivery System

David Chao

College of Business, San Francisco State University, San Francisco, CA 94132

dchao@sfsu.edu

ABSTRACT

A web document snapshot is a point-in-time capture of its code and the resulting presentation of executing the code. It is used as a way of electronically preserving historical information published in web documents enabling an organization to audit a web document's contents at a point in the past and perform business analyses with historical information recorded in it. It is also an archived copy of a web document when it is changed. This research develops a system to deliver snapshots of a web document's static and dynamic contents when it is requested. The system consists of a Database Snapshot Manager for providing database snapshots and a Web Document Snapshot Manager for providing web document snapshots. Algorithms supporting the two managers are presented.

Keywords: web document snapshot, website snapshot management

1. INTRODUCTION

A web document snapshot is the state of a web document at a point in time (snaptime).

It enables an organization to audit a web document's contents at snaptime and perform business analyses with historical information recorded in it. It is also an archived copy of a web document when it is changed [6]. Many organizations such as government are mandated to electronically archive official web documents [7] [8], a web document snapshots management system will help to meet that requirement.

However, what is the state of a web document? A web document has dual definitions. Physically, it is a text file containing code, such as HTML or XML that defines its contents and presentation. Logically, it is a web page as displayed with a browser. From a user's perspective, the rendering of a web document matters more than its source code. Considering the fact that the code of a dynamic web document may render different contents at different times, therefore, the state of a web document is a point-in-time capture of its code and the resulting presentation of executing the code.

Factors affecting the state of a web document include:

- a. Web document code: This includes the code that creates web page's static and dynamic contents. Frequently code creating dynamic contents is stored separately in a different file as in the case of Server-Side-Include and Code-Behind technologies.
- b. The state of internal resources it references: Internal resources are files managed by a web site and are available in creating the web site's contents. Typical examples of internal resources referenced by a web document that affect its rendering are style sheets, files embedded by a Server-Site Include directive, image files, script files, and databases. These internal

resources may create dynamic contents that change every time the document is opened, and they are subject to change that consequently changes the web document's rendering. Many of these files are *internal supporting files* that are created for supporting a web document and are not for publishing individually.

c. The state of external resources it references: External resources are files not managed by the web site but can be referenced in creating the web site's contents. A web document may reference external style sheets, components, web services, or databases. These external resources are also subject to change.

d. Web site host environment variables: Script code may reference a host's system variables in creating dynamic contents. A typical example is using the system clock to get the current date and time. A web document that displays the current date and time is always in a new state.

Four levels of web document snapshot Based on the four factors affecting web document's rendering, four levels of a web document's snapshot can be defined.

Level 1 snapshot: A web document snapshot is the state of web document code at snaptime. Creating level 1 snapshot enables a web site to trace the changes to the web document code over time.

Level 2 snapshot: A level 2 snapshot is a level 1 snapshot with the additional requirement that all the internal resources it references are at least level 1 snapshots at the same snaptime. These internal resources can be categorized into two groups: database and non-database files. For a database file to be a level 1 snapshot, it must be consistent with its state at the snapshot time. For a non-database file to be a level 1 snapshot, it must meet the definition above.

Level 3 snapshot: A level 3 snapshot is a level 2 snapshot with the additional requirement that all the external resources it references are at least level 2 snapshots at the snaptime.

Level 4 snapshot: A level 4 snapshot is a level 3 snapshot with the additional requirement that all the web site host's environment variables are reset to their values at snaptime.

The level 1 and level 2 snapshots are related to files that are managed by the web site, and changes to these files can be recorded. The level 3 and level 4 snapshots involve resources that are not managed by the web site and it's difficult, if not impossible, for the web site to keep track of changes to these resources. This research develops a web document snapshot management system to deliver web documents' level 1 and level 2 snapshots. Thus, the system consists of two modules: Database Snapshot Manager for maintaining database snapshots and Web Document Snapshot Manager for maintaining web document snapshots. The system is designed to deliver snapshots with any snaptime requirement, and snapshots are created only when requested. The Database Snapshot Manager is presented first.

2. DATABASE SNAPSHOT MANAGER

Database snapshot management and related materialized view management have been the topics of much research [2] [3] [4] [5]. Many schemes have been developed in maintaining database snapshots. This research adopts a generalized scheme that will meet the requirement of web document snapshots. It is presented below.

The objective of this module is to provide a database snapshot at any snaptime requested by users. This requires recording all updates in a log. The log uses time stamp to record update time, and use flags to indicate deletions and insertions where a modification is treated as the deletion of the old version followed by an insertion of the new version. With such an update log available, rolling back the current database using updates with time stamp later than the snaptime can generate a snapshot at any snaptime.

An alternative scheme is creating snapshots at different intervals. These snapshots can be *refreshed* to a new snaptime by applying updates between the old snaptime and the new snaptime. The refresh operation can go forward if the new snaptime is later than the old snaptime, otherwise it can go backward. With the availability of the two refresh operations, the optimal way is to select the snapshot whose snaptime is closest to the new snapshot, and only use updates between the snaptimes of the selected snapshot and the new snapshot. Let $UpdateLog(t_{min})$ represent the update log whose minimum time stamp value is t_{min} ; and let $UpdateLog(t_a,$

$t_b)$ represent the segment of the update log with time stamp between t_a and t_b ; assume there are n snapshots created each with snaptime t_i represented by $S(t_i)$ where i assumes values between 1 and n ; assume the new snapshot's snaptime is t_s , the algorithm $GenerateDatabaseSnapshot$ takes t_s as input and returns $S(t_s)$:

Algorithm GenerateDatabaseSnapshot:

Inputs: n snapshots, update log, and snaptime

Output: Database snapshot at snaptime

```

If  $t_s < t_{min}$  Then
  Unable to generate snapshot using this log
Else
  Select  $S(t_i)$  where  $t_i$  has the min  $|t_s - t_i|$ 
    where  $i = 1$  to  $n$ 
  If  $t_i < t_s$  Then
     $S(t_s) = S(t_i).RefreshForward(UpdateLog(t_i, t_s))$ 
  Else
     $S(t_s) = S(t_i).RefreshBackward(UpdateLog(t_i, t_s))$ 
  End if
End if

```

3. WEB DOCUMENT SNAPSHOT MANAGER

The objective of the Web Document Snapshot Manager is to generate level 2 snapshots for all internal non-database files including the supporting files. On the Internet, a Uniform Resource Locator (URL) uniquely identifies a web document. The snapshot manager is designed to let users to use a URL with a specified snaptime to retrieve the document's snapshot. An analysis of the problem is given below.

3.1 Problem analysis

At a specific point of time, a Uniform Resource Locator (URL) uniquely identifies a document on the Internet. At the web site, it translates to a physical path to the web document. Although there exists a 1:1 relationship between a URL and a web document at a given time, in the life of the document the relationship is M:M. In a document's life, it may have associated with many URLs. The web site may reorganize by changing its directory structure, hence changing the path to a document; a document may be renamed, or moved to a different directory. These changes will give the document a new URL. A URL may have associated with many documents also. It may point to a document that is different from the one it pointed to earlier. For instance, a document A may be renamed to C and another document B may be renamed to A; hence the URL originally pointing to A is now pointing to a different document. It may also happen when a web site removes a URL but later reinstates the URL for another document.

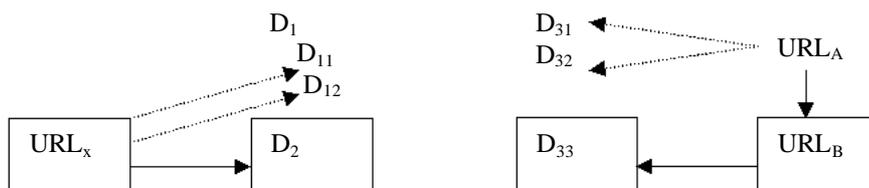


Figure 1: Relationship between URLs and web documents.

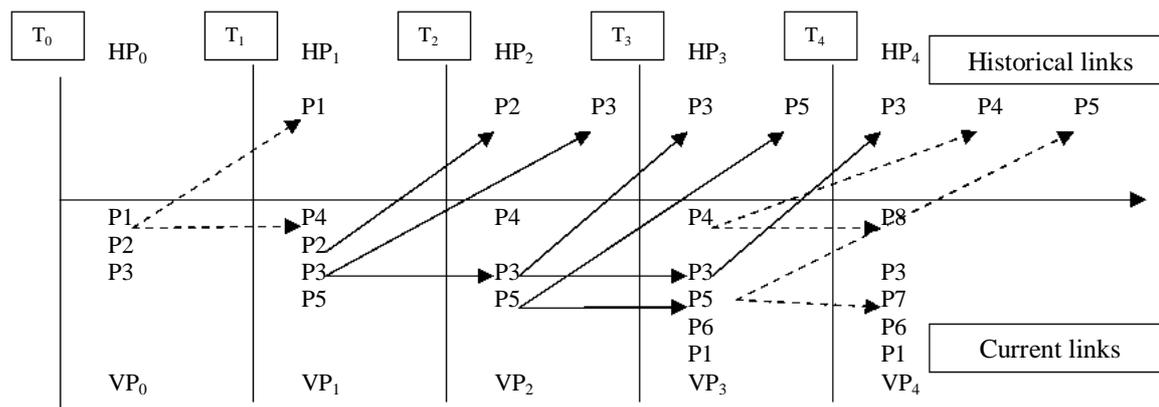


Figure 2: An example of the progression of a web site with changes since its initiation.

A web document may change its contents. Hence a URL may not only associate with many different documents, it may also associate with many versions of the same document. Figure 1 illustrates the relationship between URLs and web documents. Let D_{ij} represents document i , version j . In Figure 1a, URL_x initially points to D_{11} and D_{12} , then points to D_2 . In Figure 1b, document D_3 initially associates with URL_A , then associates with URL_B .

URLs can become invalid and users may submit invalid URLs. In Figure 1b, URL_A changes to URL_B and points to D_{33} . Is the web site able to return D_3 if users submit URL_A ? Or is it able to return D_{32} it originally pointed to? Even valid URLs don't necessarily retrieve the documents users intend for it to. In Figure 1a, a user may use URL_x saved in browser's Favorites or Bookmarks trying to retrieve D_1 . URLs invalidated by a web site due to reorganization, document removal, renaming, or relocation, plus the links to document snapshots are a web site's historical links. Tracking historical links by recording the time and the type of changes that occurred to a document and its URL enables a web site to retrieve the documents associated with the historical links.

Figure 2 illustrates the progression of a web site. T_0 is the time when the web site is initiated, and T_i denotes a point in time when its contents change. Period i is the interval of time between T_i and T_{i+1} . Assuming the path of a valid URL is unique within the web site, a valid path leads to a document currently published. In the sequel, paths and URLs are used interchangeably.

There are two sets of paths in period i : VP_i , a set of all valid paths, and HP_i , a set of paths becoming historical at T_i . Note that HP_0 is null. In Figure 2, paths that produce two dotted arrow lines such as $P1$ in period 0, and $P4$ and $P5$ in period 3 represent documents that are renamed or relocated. In this case, only the document's path has changed but the document has not changed. Therefore, there is no need to archive any documents, but it is necessary to chain the old path to the new path in order to track the change. Paths that produce two solid arrow lines such as $P3$ in period 1, and $P3$ and $P5$ in period 2 represent documents that are modified. In this case, the path is still valid but the document before the change needs to be archived as the document's snapshot. Paths that produce one solid arrow line, such as $P2$ in period 1 and $P3$ in period 3, represent documents that have been deleted. Those documents must be archived.

A path in HP_i may be repeated in HP_j , just as $P3$ is repeated in HP_2 and HP_3 . To make a historical link unique, the time the path was published is added to the path. Hence $P3$ in HP_2 is identified as $P3 + T_1$ and $P3$ in HP_3 is identified as $P3 + T_2$. Similarly, a current path may duplicate a path in the historical links, as is the case of $P1$. By adding its published time, $P1 + T_3$ is distinguishable from the historical link $P1 + T_0$. The historical links of a web site before T_i are the union of sets HP_0 to HP_{i-1} , $(HP_0 \cup HP_1 \cup \dots \cup HP_{i-1})$. Assuming T_i is the last time the web site has changed, then the web site's historical links are the unions of sets HP_0 to HP_i , and the current links are VP_i . The union of historical links and the current links represents all the paths that

are currently in use or have been used in the past. Every link in the union is uniquely identified by the value (path + path published time). This composite value is an example of a Temporal URL (discussed in the next section). For any path P_x published at time T , the temporal URL ($P_x + T$) uniquely identifies a document that is currently, or was once, associated with P_x at time T .

3.2 Solution Designs

This section explains the scheme to track the historical links of a web site. The scheme has two major components: logging the changes to web documents and archiving deleted documents including document snapshots. The log, named TemporalURLLog, is designed to keep the history of changes to web documents. It has four fields: URL, PublishDate, ExpireDate, and NewURL. The URL field records a document's path; the PublishDate records the time the document is published; the ExpireDate field records the time this URL becomes invalid; should the change create a new URL for the document, the NewURL field records the document's new URL and serves as a link to chain all log entries related to the same document together. This log has a composite key of URL + PublishDate. The value of URL + PublishDate uniquely identifies a document in a web site's history. The Archive is a directory that stores deleted files and document snapshots. Those archived files are saved in the Archive using URL + PublishDate as its file name. The log is maintained according to the log maintenance algorithm described below:

TemporalURLLog maintenance algorithm Changes to web documents are recorded in the log by the following rules:

New document: When a new document is added to the web site, a new entry is entered with its path and the time the document is published. The ExpireDate and the NewURL are set to null. Hence, log entries with a null ExpireDate are current document entries. Initially, all current documents have an entry in the log with null ExpireDate and null NewURL.

Deleted document: The log entry associated with the document is the entry with the URL equal to the document's URL with a null ExpireDate. First, it locates the entry and changes its ExpireDate to the time the document is deleted. Then, it saves the deleted document in the Archive with URL + PublishDate as its file name.

Modified document: When a document is modified, its old version becomes a snapshot and its new version is treated as a new document. The log entry associated with the document is the entry with the URL equal to the document's URL with a null ExpireDate. First, it locates the entry and changes its ExpireDate to the time

the document is modified. Then, it saves the old version in the Archive with URL + PublishDate as file name. Then, it adds a new entry with the same URL and the PublishDate is set to the time the document is modified. The ExpireDate and NewURL are set to null.

Consequently, log entries with a non-null ExpireDate and a null NewURL may be related to snapshots or deleted documents. For such an entry, if there exists another entry with the same URL and its PublishDate equals to this entry's ExpireDate then this entry is a snapshot entry and the snapshot it associates with can be retrieved from the Archive using URL + PublishDate as its file name. The snapshot is valid from the PublishDate to the ExpireDate. Otherwise, the entry is associated with a deleted document which can be retrieved from the Archive using URL + PublishDate as its file name.

Relocated or renamed page: When a document is relocated or renamed, its old URL is expired and a new URL is created. The log entry associated with the document is an entry with the URL equal to the document's URL and with a null ExpireDate. First, it locates the entry and changes its ExpireDate to the time the document is relocated or renamed and changes its NewURL field to the document's new URL. Then, it adds a new entry with the new URL and the PublishDate is set to the time the document is relocated or renamed. Hence, log entries with a non-null NewURL field help chain a document's log entries.

Using this log maintenance algorithm for the changes described in figure 2, the contents of the TemporalURLLog are shown in Figure 3, with five files in the Archive: P2T₀, P3T₀, P3T₂, P5T₁, and P3T₃. With the TemporalURLLog, a web server is able to determine that:

. A URL P2 valid between T₀ and T₁ is deleted, and the document it pointed to is in the Archive with the name P2T₀.

. A URL P3 has been modified repeatedly and is eventually deleted. All documents associated with P3 can be found in the Archive.

. An old URL P5 is now renamed to P7. It has been modified on T₃, and a copy of its snapshot can be found in the Archive with the name P5T₁.

. The log is able to determine that a historical link P1 is now renamed to P8.

. A URL P12 has never existed in the web site.

URL	PublishDate	ExpireDate	NewURL
P1	T ₀	T ₁	P4
P2	T ₀	T ₂	Null
P3	T ₀	T ₂	Null
P4	T ₁	T ₄	P8
P5	T ₁	T ₃	Null
P3	T ₂	T ₃	Null
P3	T ₃	T ₄	Null

P5	T ₃	T ₄	P7
P6	T ₃	Null	Null
P1	T ₃	Null	Null
P8	T ₄	Null	Null
P7	T ₄	Null	Null

Figure 3: The contents of TemporalURLLog based on changes described in Figure 2.

3.3 Snapshot-retrieving algorithm

This section presents an algorithm for retrieving document snapshots. The input to this algorithm is a URL P_x with a snaptime T . It assumes P_x is a current URL to simulate a scenario in which users who are viewing a current document would like to view the document's snapshot at time T . The algorithm is presented in [1] and is explained below.

It processes log entries backward starting from a document's current entry to trace back its changes in order to locate the snapshot. If the current URL's PublishDate is less than T , then the current document itself is its snapshot at T . Otherwise the backward search starts. To trace back, an entry's predecessor has one of the following properties:

1. Its URL equals the current entry's URL and its ExpireDate equals the current entry's PublishDate. This indicates the predecessor is an old version of the document and it is the document's snapshot from the predecessor's PublishDate to its ExpireDate.
2. Its NewURL equals the current entry's URL and the ExpireDate equals the current entry's PublishDate. This indicates the predecessor has been renamed or relocated.

Initially, the current document is treated as a candidate for the snapshot. Before reaching past time T , whenever an old version is found it becomes the new candidate for the snapshot. Note that an old document's life may span before its PublishDate. If a document Y is originally derived from a document X through a series of renaming or relocation, then this document's life span is from document X 's PublishDate to document Y 's ExpireDate. The search reaches the last entry when its PublishDate is before T or it no longer has a predecessor. If the last entry's PublishDate is greater than T then snapshot does not exist at T .

To illustrate, if a request $P7$ with snaptime T_2 is submitted, the backward search will pick up three entries: $(P7, T_4, \text{Null}, \text{Null})$, $(P5, T_3, T_4, P7)$, and $(P5, T_1, T_3, \text{Null})$. They show that the document associated with $P7$ has been unchanged since T_3 and was originally associated with $P5$ and renamed to $P7$ at T_4 . The document has been modified at T_3 . Therefore, the algorithm returns $\text{Archive}(P5 + T_1)$ as a snapshot. If $P7$ with snaptime T_3 is submitted, the backward search will pick up two entries: $(P7, T_4, \text{Null}, \text{Null})$ and $(P5, T_3, T_4, P7)$ and return $\text{Document}(P7)$ as a snapshot.

If a request $P8$ with snaptime T_0 is submitted, the backward search will pick up three entries: $(P8, T_4, \text{Null}, \text{Null})$, $(P4, T_1, T_4, P8)$ and $(P1, T_0, T_1, P4)$. These show that the document associated with $P8$ has been renamed at T_1 and T_4 . It has not been modified since T_0 . The algorithm will return $\text{Document}(P8)$ as a snapshot.

4. WEB DOCUMENT SNAPSHOTS DELIVERY METHODS

This section discusses schemes for users to submit a request for a snapshot. Such requests must include the document's URL and snaptime required by the snapshot-retrieving algorithm. Three schemes for submitting a request are proposed below:

Using TemporalURL A temporal URL is a URL submitted with temporal requirements of which the documents associated with the URL must meet [1]. A typical way to submit additional information with a URL is through query strings. A query string is a set of name=value pairs appended to a URL. It is created by adding a question mark (?) immediately after a URL followed by name=value pairs separated by ampersands (&). For instance, a Snaptime query string may be defined as: $P_x?\text{Snaptime}=T$ to retrieve the web document's snapshot at time T .

Problems with using TemporalURL are: (1) implementing this scheme requires the web server be changed to distinguish between requests for regular documents and requests for snapshots. This causes interruption for non-snapshot requests. (2) Users must be informed of such service and be trained to use temporal URL.

Using a snapshot management site This is a site designed to handle snapshot requests. Its objectives are: (1) educating users about the web document snapshot systems and (2) providing interface to enter request for snapshots. The interface can be a form with textboxes to enter web document's URL and snaptime. The snapshot-retrieving algorithm can be implemented with the server-side program that handles the form.

The advantage of this approach is that the web server does not need to change its behavior because of snapshots.

Using web services A web service is an application logic accessible via the Internet [9]. It is defined as a function that takes inputs from its arguments to perform a task. It uses standard Internet protocols such as HTTP and SOAP for communication. Inputs to a web service can be submitted with HTTP GET, or HTTP POST, or enclosed in a SOAP Envelop encoded in XML. The output from the web service is typically returned in a SOAP Envelop encoded in XML. The inputs and outputs are all in text format. A web service can be

accessed from a browser. In this mode, it will provide an interface page with textboxes to enter inputs to the service along with an information page explaining input data type. It can also be integrated in an application where it is referenced as a component. In this mode, it is used as a function where the application provides inputs through its arguments.

The snapshot-retrieving algorithm can be implemented as a web service that takes URL and snaptime as its inputs and returns the document snapshot as output. Because web services communicate in text, the snapshot returns to the user will be in text format. This implies that it is the web document's code that is sent to the user, not the web page as displayed with a browser. Hence any server-side program in the web document is unable to run. Therefore, it is a level 1 snapshot.

5. SUMMARY

Web document snapshots are useful for archiving or performing business analysis with historical data. This paper defined four levels of snapshot contents where the level 1 and level 2 snapshots involve resources managed by the web site, and level 3 and level 4 involve resources not managed by the web site. This paper proposes a web document snapshot management system designed to deliver level 2 snapshots. The system consists of a Database Snapshot Manager for providing database snapshots and a Web Document Snapshot Manager for providing web document snapshots. Algorithms supporting the two managers are presented. These algorithms use logs with time stamps to record changes to the database and web documents. With the logs, snapshots can be generated dynamically when requested. Three types of interface between the web site and users regarding snapshot delivery are proposed: temporal URL, snapshot handling site, and web service. With the proposed snapshot management system, a web site is able to deliver snapshots with dynamic contents.

REFERENCES

- [1] Chao, D. (2003). Tracking a Web Site's Historical Links with Temporal URLs. Proceedings of the 3rd International Conference on Electronic Business, Singapore, 2003.
- [2] Chao, D., Diehr, G., & Saharia, A. (1996) Maintaining Join-based Remote Snapshots Using Relevant Logging. Proceedings of the Workshop on Materialized Views, ACM SIGMOD, Montreal, Canada, 1996.
- [3] Chien, S., Tsotras, V., & Zaniolo, C. (2001) Efficient Management of Multiversion Documents by Object Referencing. Proceedings of 13th International Conference on Very large Data Bases, 2001.
- [4] Labrinidis, A. & Roussopoulos, N. (2000). Webview Materialization. ACM SIGMOD International Conference on Management of Data, May 14-19, 2000.
- [5] Marian, A., Gregory Cobena, S., & Mignet, L (2001) Change-Centric management of Versions in an XML Warehouse. Proceedings of the 27th VLDB Conference, Rome, Italy, 2001.
- [6] Smithsonian Institution, "Archiving Smithsonian Websites", <http://www.si.edu/archives/archives/websitepilot.html#intro> (May 2003)
- [7] U.S. National Archives & Record Administration, "Federal Web Site Snapshot Information", http://www.archives.gov/records_management/web_site_snapshot/snapshot.html,
- [8] Utah State Government, "Draft Electronic Records Policy", <http://www.archives.state.ut.us/recmanag/electronicpolicy.htm>, Feb. 2002
- [9] W3C World Wide Web Consortium, "Web Services Architecture", <http://www.w3.org/TR/ws-arch>, Feb. 2004

Using Intelligent Agents to Build E-Business Software

Adrien Coyette, Manuel Kolp, Stéphane Faulkner

Information Systems Research Unit, University of Louvain,
1 Place des Doyens, 1348 Louvain-la-Neuve, Belgium
{coyette, kolp, faulkner}@isys.ucl.ac.be

ABSTRACT

Agent architectures are gaining popularity for building open, distributed, and evolving software required by e-commerce applications. Unfortunately, despite considerable work in software architecture during the last decade, few research efforts have aimed at truly defining patterns and languages for agent architectural design. This paper proposes a modern approach based on organizational structures and architectural description languages to define and specify agent architectures notably in the case of e-commerce system design.

Keywords: Agent Systems, Architectural Description Language, Organizational Styles, E-commerce Application

1. INTRODUCTION

The meteoric rise of Internet and World-Wide-Web technologies has created overnight new application areas for enterprise software, including e-commerce applications. These areas demand software that is robust, can operate within a wide range of environments, and can evolve over time to cope with changing requirements. Moreover, such software has to be highly customisable to meet the needs of a wide range of users and sufficiently secure to protect personal data and other assets on behalf of its stakeholders.

Not surprisingly, researchers are looking for new software designs that cope with such requirements. One promising source of ideas for designing such e-commerce software is the area of agent architectures. They appear to be more flexible, modular and robust than traditional including object-oriented ones. They tend to be open and dynamic in the sense they exist in a changing organizational and operational environment where new components can be added, modified or removed at any time.

To cope with the ever-increasing complexity of the design of software architecture, architectural design has received through the last decade increasing attention as an important field of software engineering. Practitioners have come to realize that getting an architecture right is a critical success factor for system life-cycle and have recognized the value of making explicit architectural descriptions and choices in the development of new software.

To this end, a number of architectural description languages (ADL) and architectural styles [4] have been proposed for representing and analyzing architectural designs. An architectural description language provides a concrete syntax for specifying architectural abstractions in a descriptive notation while an architectural style constitutes an intellectually manageable abstraction of system structure that

describes how system components interact and work together.

Unfortunately, despite this considerable work [9], few research efforts have aimed at truly defining styles and description languages for agent architectural design. To fill this gap, we have defined, in the SKwyRL project (<http://www.isys.ucl.ac.be/skwyrl/>), architectural styles for agent systems based on an organizational perspective [2] and have proposed in [3] SKwyRL-ADL, an agent architectural description language. This paper continues and integrates this research: it focuses on an agent perspective for designing and specifying e-commerce software architecture based on organizational styles and SKwyRL-ADL. The structure-in-5 organizational style will be instantiated to design the architecture of the system and the specifications will be expressed in a formal way with SKwyRL-ADL. The rest of the paper is organized as follows. Section 2 introduces some perspectives of SKwyRL insisting on the BDI model, our ADL and organizational styles. Section 3 describes our agent oriented approach on e-commerce system development, including the design of the global architecture with organizational styles, its formal specification with SKwyRL-ADL and the corresponding implementation on an agent-oriented platform. Finally, Section 4 concludes the research..

2. ADLAND STYLES IN SKWYRL

We have detailed in the SKwyRL project an agent ADL called SKwyRL-ADL [3] that proposes a set of abstractions that are fundamental to the description and specification of agent architectures based on the BDI (Belief-Desire-Intention) agent model. To help the reader to understand our ADL specification in the rest of the paper, we briefly present the main elements of SKwyRL-ADL including the BDI agent model. SKwyRL-ADL is composed of two sub-models which operate at two different levels of abstraction: internal and global. The internal model captures the states of

an agent and its potential behavior. The global model describes the interaction among agents that compose the agent architecture. We will also introduce organizational styles through the description of one of them, the structure-in-5, that will be used later on in the paper.

2.1 The BDI Agent Model

An agent defines a system entity, situated in some environment that is capable of flexible autonomous action in order to meet its design objective [11].

An agent can be useful as a stand-alone entity that delegates particular tasks on behalf of a user. However, in the overwhelming majority of cases, agents exist in an environment that contains other agents. Such environment is a agent system that can be defined as an organization composed of autonomous and proactive agents that interact with each other to achieve common or private goals [6].

In order to reason about themselves and act in an autonomous way, agents are usually built on rationale models and reasoning strategies that have roots in various disciplines including artificial intelligence, cognitive science, psychology or philosophy. An exhaustive evaluation of these models would be out of the scope of this paper or even this research work. A simple yet powerful and mature model coming from cognitive science and philosophy that has received a great deal of attention, notably in artificial intelligence, is the Belief-Desire-Intention (BDI) model [1]. This approach has been intensively used to study the design rationale of agents and is proposed as a keystone model in numerous agent-oriented development environments such as JACK [5]. The main concepts of the BDI agent model are in addition to the notion of agent itself we have just explained:

- Beliefs that represent the informational state of a BDI agent, that is, what it knows about itself and the world;
- Desires (or goals) that are its motivational state, that is, what the agent is trying to achieve;
- Intentions that represent the deliberative state of the agent, that is, which plans the agent has chosen for possible execution.

2.2 Internal Model

Figure 1 illustrates the main entities and relationships of the internal model of SKwyRL-ADL. The agent needs knowledge about the environment in order to reach decisions. Knowledge is contained in agents in the form of one of many knowledge bases. A Knowledge base consists of a set of beliefs that the agent has about the environment and a set of goals that it pursues. A belief is a finite set of objects, things with individual identities and properties, that represents a view of the current environment states of an agent. However, beliefs about the current state of the environment are not always

enough to decide what to do. In other words, as well as a current state description, the agent needs some sort of goal information, which describes an environment state that are (not) desirable.

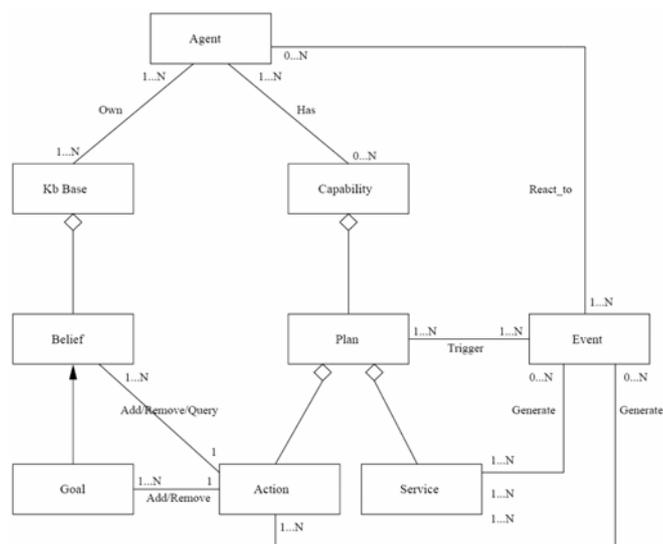


Figure 1: Conceptual representation of the internal model

The intentional behavior of an agent is represented by their capabilities to react to events. An event is generated either by an action that modifies beliefs or adds new goals, or by services provided from another agent. Note that these services are represented in the global model because they involve interaction among agents that compose the agent system.

An event may invoke (trigger) one or more plans; the agent commits to execute one of them, that is, it becomes intention. A plan defines the sequence of action to be chosen by the agent to accomplish a task or achieve a goal. An action can query or change the beliefs, generate new events or submit new goals.

2.3 Global Model

Figure 2 conceptualizes the global model which describes the interaction among agents that compose the agent system. Configurations are the central concept of architectural design, consisting of an interconnected set of agents. The topology of a configuration is defined by a set of bindings between provided and required services.

An agent interacts with its environment through an interface composed of sensors and effectors. An effector provides to the environment a set of services. Then, a sensor requires a set of services from the environment. A service is an action involving an interaction among agents.

The whole agent system is specified with an architecture which contains a set of configurations. An architecture

represents agents by one or more detailed, lower-level configuration descriptions.

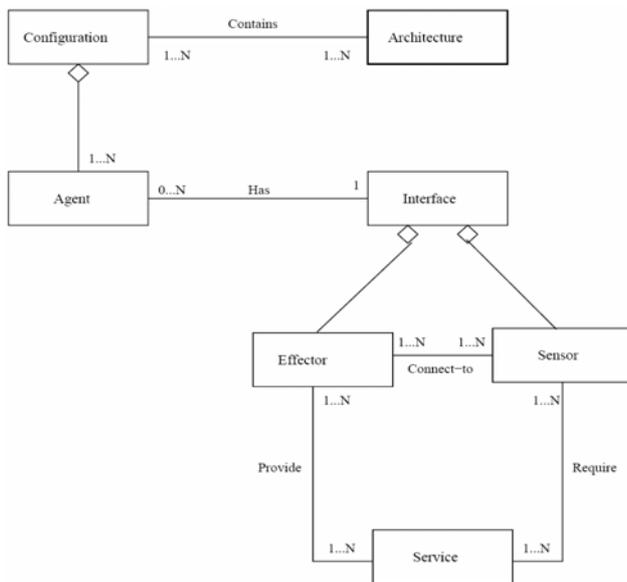


Figure 2: Conceptual Representation of the Global Model

2.4 Agent Architectural Styles

A key aspect to conduct architectural design in SKwyRL is the specification and use of organizational styles [2, 6]. These are socially-based design alternative inspired from models and concepts from organizational theories that analyze the structure and design of real-world human organizations.

For instance, the agent architecture we propose in Figure 3 has been designed following the structure-in-5 organizational style detailed in [2]. In a few words, the structure-in-5 style is a meta-structure that defines an organizational architecture that as an aggregate of five sub-structures, as described by Mintzberg [8]. It consists of five typical strategic and logistic components found in many organizations. At the base level one finds the Operational Core where the basic tasks and operations -- the input, processing, output and direct support procedures associated with running the system -- are carried out. At the top lies the Apex composed of strategic executive components. Below it, sit the control/standardization, management components and logistics: Coordination, Middle Agency and Support, respectively.

3. AGENT ARCHITECTURE FOR E-COMMERCE SYSTEM

E-Media (<http://www.isys.ucl.ac.be/skwyrl/emedial>) is a typical business-to-consumer application we have developed using the architectural concepts explained in Section 2. The application offers an e-commerce architecture supporting the creation of information

sources that facilitate the on-line transaction of products, services, and payments resulting in an effective and efficient interaction among sellers, buyers and intermediaries.

This section describes how we have applied the structure-in-5 style to design the architecture of E-media and used SKwyRL-ADL to formally specify each architectural aspect (belief, goal, plan, action, interface, configuration, service ...) of the application. Based on this architectural specification, we have implemented the application using JACK, a JAVA agent-oriented development environment.

3.1 E-Media Architecture

E-Media includes the following features:

- An on-line web interface allows customers to examine the items in the E-Media catalogue, and place orders;
- Customers can search the on-line store by either browsing the catalogue or querying the item database. An online search engine allows customers to search title, author/artist and description fields through keywords or full-text search;
- If an item is not available in the catalogue, the customer has the option to order it;
- Internet communications are supported;
- On-line financial transactions including credit card and anonymity are protected;
- All web information (e.g., product and customer turnover, sales average, ...) of strategic importance is recorded for monthly or on-demand statistical analysis;
- Based of this statistical and strategic information, the system permanently manages and adapts the stock, pricing and promotions policy. For example, for each product, the system can decide to increase or decrease stocks or profit margins. It can also adapt the customer on-line interface with new product promotions.

Figure 3 models the architecture of E-Media using the i* model [12] following the structure-in-5 organizational style we have introduced in Section 2.

i* is a graph, where each node represents an actor (or system component) and each link between two actors indicates that one actor depends on the other for some goal to be attained. A dependency describes an "agreement" (called dependum) between two actors: the depender and the dependee. The depender is the depending actor, and the dependee, the actor who is depended upon. The type of the dependency describes the nature of the agreement. Goal dependencies represent delegation of responsibility for fulfilling a goal; softgoal dependencies are similar to goal dependencies, but their fulfilment cannot be defined precisely; task dependencies are used in situations where the dependee is required. As show in Figure 3, actors are represented as circles; dependums – goals, softgoals, tasks and resources – are respectively represented as ovals, clouds, hexagons and rectangles;

dependencies have the form *dependeur* → *dependum* → *dependee*.

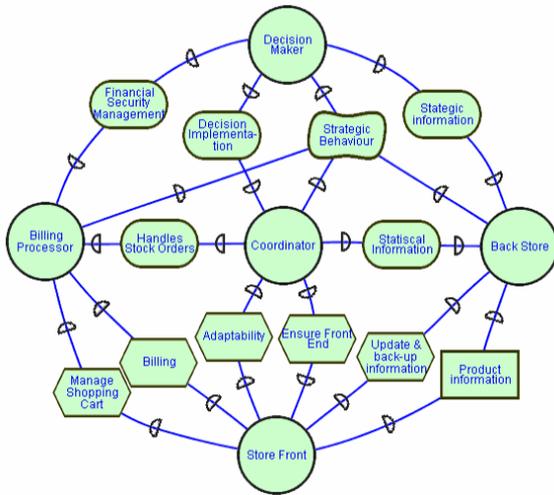


Figure3: The E-Media Architecture in Structure-in-5

The Store Front plays the role of the structure-in-5's Operational Core. It interacts with customers and provides them with a usable front-end web application for consulting, searching and shopping media items.

The Back Store constitutes the structure-in-5's Support component. It manages the product database and communicates to the Store Front relevant product information. It stores and backs up all web information about customers, products and sales to be able to produce statistical information (e.g., analyses, average charts and turnover reports). Such kind of information is computed either for a predefined product (when the Coordinator asks it) or on a monthly basis for every product. Based on this monthly statistical information, it provides also the Decision Maker with strategic information (e.g., sales increase or decrease, performance charts, best sales, sales prevision, ...).

The Billing Processor plays the role of the structure-in-5's Technostructure in handling customer orders and bills. To this end, it provides the customer with on-line shopping cart capabilities. It also ensures the secure management of financial transactions for the Decision Maker. Finally, it handles, under the responsibility of the Coordinator component, stock orders to avoid shortages or congestions.

As the structure-in-5's Middle Agency, the Coordinator assumes the central position of the architecture. It is responsible to implements strategic decisions for the Decision Maker (Strategic Apex). It supervises and coordinates the activities of the Billing Processor (initiating the stock and pricing policy), the Front Store (adapting the front end interface with new promotions and recommendations) and the Back Store

(parametrizing statistical computing) ensuring that the system fulfills its mission in an effective way.

Finally, the Decision Maker assumes the Strategic Apex role of the structure-in-5. It defines the Strategic Behavior (e.g., sales and turnover, product visibility, hits, ...) of the system ensuring that objectives and responsibilities delegated to the Billing Processor, Coordinator and Back Store are consistent with respect to their capabilities.

3.2 E-Media Formal Specification

The architecture described in Figure 3 gives an organizational representation of the system-to-be including relevant actors and their respective goals, tasks and resource inter-dependencies. This model can serve as a basis to understand and discuss the assignment of system functionalities but it is not adequate to provide a precise specification of the system details. As introduced in Section 2, SKwyRL-ADL provides a finite set of formal agent-oriented constructors that allow to detail in a formal and consistent way the software architecture as well as its agent components and their behaviors.

```

Agent:{Back-Store
Interface
Effector[provide(strategic_info)]
Effector[provide(statistical_info)]
Effector[provide(product_info)]
Effector[provide(back-up)]
Sensor[require(strategic_behavior)]
KnowledgeBase:
Product_KB      Statistical_KB
BS_System_KB   BS_Customer_KB
Capabilities:
Statistical_CP  Strategic_CP
Data_Management_CP
}

```

Figure 4: Agent Description of the Back-Store

Figure 4 shows a high-level formal description of the Back-Store agent. Three aspects of this agent component are of concern here: the interface representing the interactions in which the agent will participate, the knowledge base defining the agent knowledge capacity and the capabilities defining agent behaviors.

SkwyRL-ADL allows to work at different levels of architectural abstractions (i.e., different views of the system architecture) to encapsulate different components of the system in independent hierarchical descriptions. For instance, in Figure 4 the Back Store agent has three knowledge bases (KB) and three capabilities (CP), but the description level chosen here does not specify the details of the beliefs composing the KB or the plans and events composing each capability.

The rest of the section focuses on the Back Store agent to give an example of a refinement specification with

our ADL for each of the three aspects of the agent: interface, knowledge bases and capabilities.

Interface. The agent interface consists of a number of effectors and sensors for the agent. Each of them represents an action in which the agent will participate. Each effector provides a service that is available to other agents, and each sensor requires a service provided by another agent. The correspondence between a required and a provided service defines an interaction. For example, the Back Store provides the `statistical_info` service that the Coordinator requires.

Such interface definition points two aspects of an agent. Firstly, it indicates the expectations the agent has about the agents with which it interacts. Secondly, it reveals that the interaction relationships are a central issue of the architectural description. Such relationships are not only part of the specification of the agent behavior but reflect the potential patterns of communication that characterize the ways the system reason about itself.

```

Service: {Ask(statistical_info)
sender: Coordinator
parameters: (tw: TimeWindows), (id: Id_product)
reply_with: to: Turnover v sl: Sales
receiver: Back-Store
Effect:Add(Statistical_KB, Achieve (stat("today", "on_product")
})

```

Figure 5: A Service Specification

The specification of a service is given in Figure 5. Each provided or required service is detailed by specifying the sender agent that initiates the service, the set of receiver agents that interact with the sender, the reply-with and content statements that define the information about which the service expresses an interaction and optionally a set of parameters to define the information required to execute the service. Like the parameters the reply-with information is represented with a belief or a set of terms (e.g., function, constant or variable).

Knowledge Bases. A knowledge base (KB) is specified with a name, a body and a type. The name identifies the KB whenever an agent wants to query or modify them (add or remove a belief). The body represents a set of beliefs in the manner of a relational database schema. It describes the beliefs the agent may have in terms of fields. When the agent acquires a new belief, values for each of its fields are specified and the belief is added to the appropriate KB as a new tuple. The KB type describes the kind of formal knowledge used by the agent. A Closed world assumes that the agent is operating in a world where every tuple it can express is included in a KB at all times as being true or false. Inversely, in an open world KB, any tuple not included as true or false is assumed to be unknown. Figure 6 gives the specification the Product_KB:

```

KnowledgeBase: {Product_KB
Kb_body:

```

```

product(Id_Prod, Title, Class, Description, Price)
purchase(Id_Card, Id_Prod, Date, Quantity, Pay_Means)
audio(Id_Prod, Artist(+), Composer(+))
book(Id_Prod, ISBN, ISNN, Author(+), Publisher)
dvd(Id_Prod, Actor(+), Realizator(+))
...
kb-type: closed_world
}

```

Figure 6: A Knowledge Base Specification

Capabilities formalize the behavioral elements of an agent. They are composed of plans and events that together define the agent's abilities. They can also be composed of sub-capabilities that can be combined to provide complex behavior.

```

Capability: {Data_Management_CP
CP_body:
Plan Send_Product_Data
Plan Backup_Database
Plan Check_Identification
Plan Interface_Custom_Id
endEvent update_interface
PostEvent backup
}

```

Figure 7: Capability Specification

Figure 7 shows the Data_Management capability of the Back-Store agent. The body contains the plans the capability can execute and the events it can post to be handled by other plans or it can send to other agents. For example, the Data_Management capability is composed of four plans: Send_Product_Data is used to send product information to the Store-Front, Backup_Database backs up all information about customers, products and sales in order to provide statistical information, Interface_Custom_Id allows user registration and Check_Identification checks the customer login/password when an order is made.

A plan defines the sequence of actions and/or services (i.e., actions that involve interaction with other agents) the agent selects to accomplish a task or achieve a goal. A plan consists of:

- an invocation condition detailing the circumstances, in terms of beliefs or goals, that cause the plan to be triggered;
- an optional context that defines the preconditions of the plan, i.e., what must be believed by the agent for a plan to be selected for execution;
- the plan body, that specifies either the sequence of formulae that the agent needs to perform, a formula being either an action or a service to be executed; an end state that defines the post-conditions under which the plan succeeds;
- and optionally a set of services or actions that specify what happens when a plan fails or succeeds.

Configuration To describe the complete topology of the system architecture, the agents of an architectural description are combined into a SkwyRL configuration.

Instances of each agent or service that appear in the configuration must be identified with an explicit and

unique name. The configuration also describes the collaborations (i.e., which agent participates in which interaction) through a one-to-many mapping between provided and required service instances.

3.3 E-Media Implementation

Based on the structure-in-5 architecture described in Section 3.1 and the formal SKwyRL-ADL architectural specification overviewed in Section 3.2, the E-Media application has been implemented with JACK (JACK), a BDI agent-oriented development environment for JAVA. We briefly describe the E-Media implementation to illustrate the role of the agents and their interactions.



Figure 8: E-media Main Interface

When an on-line customer gets connected to E-media, an instance of the Front-Store is created to display the interface depicted in Figure 8. It allows the new coming user to register (1). The Back-Store handles the information provided by the user and checks its validity (2). If the access is granted, the user can purchase products on E-Media by adding catalogue items to the shopping cart (4) managed by the Billing-Processor. At any time the user can use the navigation-bar (3) to switch from one section of the website to another. Promotions (5) and best sales (6) are part of the strategic behaviour objective. The promotions policy is initiated by the Decision-Maker based on the strategic information provided by the Back-Store. The Coordinator chooses the best promotions and consequently adapts the Store Front layout. The Coordinator acts similarly for the best sales: the Back-Store computes the five best sellers and the Coordinator accordingly updates the Store-Front.

6. CONCLUSION

Nowadays, software engineering for new enterprise application domains such as e-Business is forced to build up open systems able to cope with distributed, heterogeneous, and dynamic information issues. Most of these software systems exist in a changing organizational and operational environment where new components can be added, modified or removed at any

time. For these reasons and more, agent architectures are gaining popularity in that they do allow dynamic and evolving structures which can change at run-time.

Architectural design has received considerable attention for the past decade which has resulted in a collection of well-understood architectural styles and formal architectural description languages. Unfortunately, this work has focuses on object-oriented rather than agent-oriented systems. This paper has described an approach based on organizational styles and an agent architectural description language we have defined to design agent architectures in the context of e-commerce system engineering. The paper has proposed a validation of the framework: it has been applied to develop E-Media, an e-commerce platform implemented on the JACK agent environment.

REFERENCES

- [1] Bratman, M. E., *Intention, Plans and Practical Reason*. Harvard University Press, 1987.
- [2] Do, T. T., Faulkner, S. and Kolp, M., *organizational Multi-Agent Architectures for Information Systems*. in Proc. of the 5th Int. Conf. on Enterprise Information Systems (ICEIS 2003), Angers, France, April 2003.
- [3] Faulkner, S. and Kolp, M., *Towards an Agent Architectural Description Language for Information Systems*. Proc. of the 5th Int. Conf. on Enterprise Information Systems (ICEIS 03), Angers, France, April 2003.
- [4] Garlan, D., Allen, R. and Ockerbloom, J., *Exploiting Style in Architectural Design Environments*. In Proc. of SIGSOFT'94: Foundations of Software Engineering, New Orleans, USA, Dec. 1994.
- [5] <http://www.agent-software.com/>.
- [6] Kolp, M., Giorgini, P. and Mylopoulos, J., *An Organizational Perspective on Multi-agent Architectures*. In Proc. of the 8th Int. Workshop on Agent Theories, architectures, and languages, ATAL'01, Seattle, USA, Aug. 2001.
- [7] Louis, R., 1999. *Software agents activities*. In ICEIS'99, 1st International Conference on Enterprise Information Systems. ICEIS Press.
- [8] Mintzberg, H., *Structure in fives: designing effective organizations*. Prentice-Hall, 1992.
- [9] Shaw, M. and Garlan, D., *Software Architecture: Perspectives on an Emerging Discipline*, Prentice Hall, 1996.
- [10] Smith, J., 1998. *The book, The publishing company*. London, 2nd edition.
- [11] Wooldridge, M. and Jennings, N.R., editors. *Special Issue on Intelligent Agents and Multi-Agent Systems*. Applied Artificial Intelligence Journal. Vol. 9(4), 1996.
- [12] Yu, E., *Modeling Strategic Relationships for Process Reengineering*, Ph.D. thesis, Department of Computer Science, University of Toronto, Canada, 1995.

What Data Is Necessary To Data Mine For Knowledge?

Paul Alpar

School of Business and Economics, Philipps University Marburg,
 Universitätsstr. 24 35032 Marburg, Germany
 Phone: ++49-6421-2823703, Fax: ++49-6421-2826554
 alpar@wiwi.uni-marburg.de

ABSTRACT

The data that most organizations possess is considered a valuable asset. To really benefit from it, organizations must be able to analyze these data efficiently and effectively. This activity is nowadays referred to as business intelligence and one class of algorithms used within this activity are called data mining methods. There are some famous success stories about knowledge discovered via data mining but there is also a lot of disappointment so far. The paper argues that one of the reasons for some failures to produce better results with data mining is the reliance on transactional and master data only. It points to other types of data that can enrich transactional data and help in this way to produce more interesting and more rewarding data patterns.

Keywords: knowledge discovery, data mining, data warehouse, business intelligence

1. INTRODUCTION

The computerization of most (business) processes has led to a dramatic increase in transactional data. One of the latest reasons for a revolutionary increase in transactional data is the wide-spread use of the world-wide web where each request for information is recorded in a log file. Transactional data always contain information on *what* happened and often but not always *who* the actor was. For example, in web log mining it is often not known who the web site visitor was. If we know the actor, we can combine transactional data with master data if they exist. Practitioners often misinterpret the most cited process model (see Figure 1) in such a way that transactional (and master) data are sufficient to perform knowledge discovery. A typical statement is "Data mining can help companies extract additional value from piles of accumulated customer transaction data"[7]. This is not wrong but transactional data are often not sufficient to perform successful knowledge discovery as will be shown below. Many companies first build a data warehouse, primarily for regular reporting and ad-hoc queries, and then they start thinking which more complex data analyses to perform. Building a data warehouse usually comprises merging transactional and master data from several applications and sources, a tedious task that takes a lot of resources and time [3]. If additional data are needed later for data mining they may not be available anymore or the management is reluctant to restart the process of building the data warehouse. So data mining is only performed with the data that have been collected before independently of any knowledge discovery objectives.

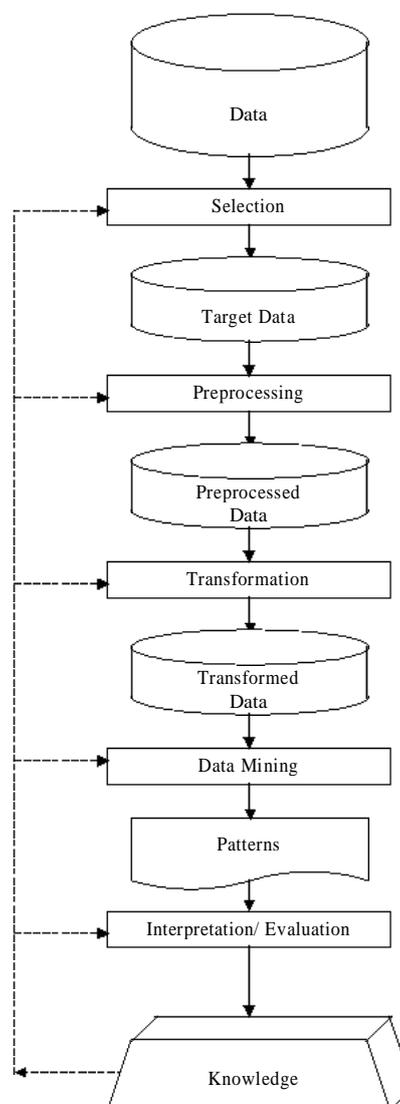


Figure 1: Knowledge Discovery Process [2]

Following such an approach, the results of the data mining process are often disappointing; they are not

interpretable, not interesting, or not actionable. Other authors suggest, therefore, to enrich data during the preprocessing step [7] or they propose an alternative process model (see Figure 2).

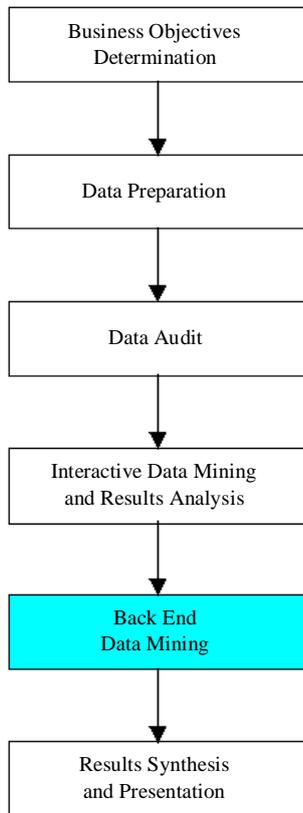


Figure 2: Knowledge Discovery Steps [4]

It is especially the step of “back end data mining” that gives rise to the hope that this process will lead to better results. According to Hirji [4] “Back end data mining involves data enrichment and additional data mining algorithm execution (...) in order to increase the dimensionality of the data mining data set with third-party demographic data (...)”

Our view, based on carrying out a number of data mining projects, is that additional data, not just demographic data, may be needed at collection time, in the selection step, and in the preprocessing step. The collection or creation of these data is based on knowledge that is sometimes loosely referred to as background, domain, or application-specific knowledge. It enables an analyst to collect, create, or select data that are relevant to the analysis task without being an explicit part of a transaction. To be able to initiate the necessary activities in due time, the analyst should roughly know ahead what s/he will be mining for. A process model proposed by SPSS, a statistical software vendor, in collaboration with some industry representatives, called CRISP-DM (Cross Industry Standard Process for Data Mining), explicitly requires to collect data only after the business and data mining objectives have been defined (see Figure 3). Further, it suggests constructing and integrating data, if needed, during the phase of data preparation. This model is, however, not specific about why and what type of new data should be generated via construction or integration.

We try to determine various types of data that may play a role in a data mining task without being part of transactional or master data. The importance of such data is demonstrated through examples from specific projects that we have conducted in the last years. The data types have been determined inductively through action research [6] so they are biased by the kind of projects undertaken. The projects cover several industries and almost all widely used data mining methods (decision trees, artificial neural networks, association rules, case-based reasoning, various clustering other statistical methods, scoring). However, it may be that mining in scientific data would reveal that additional or other data types are important.

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objects	Collect Initial Data	Select Data	Select Modeling Technique	Evaluate Results	Plan Deployment
Assess Situation	Describe Data	Clean Data	Generate Test Design	Review Process	Plan Monitoring and Maintenance
Determine Data Mining Goals	Explore Data	Construct Data	Build Model	Determine Next Steps	Produce Final Report
Produce Project Plan	Verify Data Quality	Integrate Data	Assess Model		Review Project
		Format Data			

Figure 3: CRISP-DM Reference Model [1]

2. CIRCUMSTANTIAL DATA

Circumstantial data relate to the conditions when the transaction occurred. It is usually easier to create or collect such data at the time that transactional data are collected than later, perhaps after months or years, when data mining takes place. For example, to get detailed

regional weather data by the hour and minute for each transaction after several months or years might be possible but it is certainly more difficult and expensive than recording the data when a transaction occurs. Table 1 shows an example where the task was to segment customers of a telecommunication product.

Table 1: Enriched call data

Col No	Column Name	Format	Description
...
5	from_connect_tm
6	from_weekend_ind	numeric1	Set if day called at origin = Sunday or Saturday
7	to_country_cd
8	to_connect_dt
9	to_connect_tm
10	to_weekend_ind	numeric1	Set if day called at destination = Sunday or Saturday
11	vacation_ind	numeric1	Set if day called at origin in {24.12 - 31.12 or 1.7 -31.8}
12	days_since_last_call_qty	numeric	No of days since customer called last
13	destination_typ
...

The basic call data (numbers of the caller and the person called as well as the time) were immediately augmented by data that indicate whether the call took place on the weekend (column 6), during vacation time (column 11), and a number of other indicators. Some of these played an important role during the clustering of customer groups. As can be seen from column 11, such data can be soft in the sense that they reflect an assumption or perception. In this case, it is assumed that the caller is on vacation if the call was made in the indicated time period.

Circumstantial data can be also important in the selection step. In another application, the task was to develop a scoring model with which customers of a financial institution can be targeted for specific investment products [5]. One of the variables that played an important role was the value of the customer's stock portfolio. This obviously can fluctuate considerably. To control for the fluctuations of the stock market, data about the German stocks index DAX were used in such a way that customer data were selected for training purposes from a period of time when the DAX was relatively stable (see Figure 4). In this case, this was the period between January 2, 2001 and June 29, 2001.

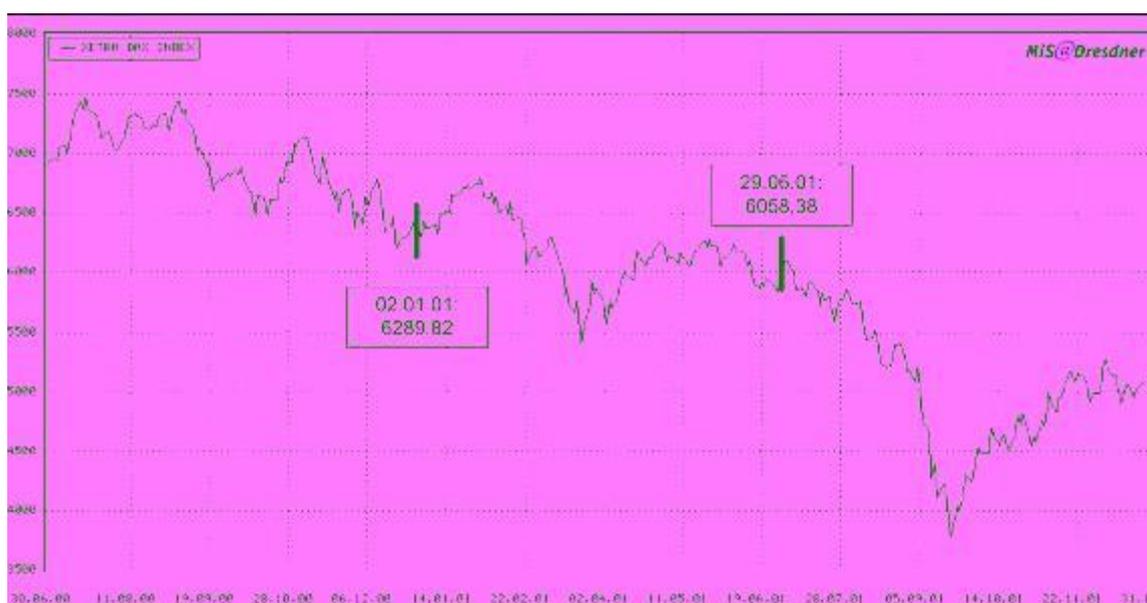


Figure 4: Run of the DAX between April 20, 2000 and December 31, 2001

3. STRUCTURAL DATA

While in some cases a simple categorization of data can be helpful, in many cases the knowledge of more complex relationships among objects involved in transactions is needed. For example, in the analysis of purchase baskets in a supermarket the scanner data allow to mine for associations at the level of universal product codes. However, given that many products from the same brand and kind come in different packages, it may be meaningful to summarize them as a pre-processing step. Sometimes, further aggregations along product categories (across brands) are needed to discover association rules. Therefore, the structure of product groups needs to be applied to transactional data before data mining can be performed.

4. QUALIFICATION DATA

Qualification data describe more precisely what has happened. Transactional data contain the facts, qualification data add meaning to these facts. In the above mentioned example, segmentation of customers of a telecommunication product, each transaction contains the origination and destination phone number. Through a comparison of the destination number with the customer's home and office numbers, which are usually recorded in the master data, each call was qualified to be a call *home*, to the *office*, to the *home city*, to the *home country*, or *other*. Another comparison of the origination and the destination numbers qualified the calls as being *in-country* or *out-of-country* calls. It turned out that one of these variables played a role in forming the customer clusters. Another example is visits to a web site. They can be relatively easily distinguished by the number of web pages requested, the visit length, or the action taken (e. g., a registration, an information request, or a purchase). However, if visitors usually do not take an action at a web site, it may be helpful to characterize each visit before further analyzing the visiting behavior. One simple qualification using cookie technology can be the distinction between *first visit* and *repeat visit*.

Another example relates to a set of purchase data from mail orders. Some of the customers acted as sales people in the sense that they were also placing collective orders for several customers. Their pay-off was a bonus on the accumulated sales amount. The mail orders did not contain any information whether an order was placed for one or several customers. The goal of the analysis was to search for association rules among products in a sales transaction (basket analysis). Collective orders would distort the results since they contain several transactions. Therefore, the mail orders were qualified as collective orders if the total order value

exceeded a threshold value. These purchases were then excluded from further analysis.

The first example shows that transactions can be qualified based on facts while the second one shows that, sometimes, they must be based on "educated guesses."

5. NON-TRANSACTIONAL TRACKING DATA

Non-transactional tracking data record the behavior of objects of interest, often customers, even when no transactions occur. When a transaction occurs transaction processing systems carry out necessary actions including the update of stock values (e.g., customer account balance or number of seats available on a plane). However, the lack of activity of an object can also be an important information worth recording. As an example, we look at data of a direct order company. The orders come via phone, mail, or internet. In each of the cases, a quick credit assessment takes place. Most cases are decided automatically by a program that decides whether to deliver the merchandise to the customer and what kind of payment to accept. This process is referred to as *behavioral scoring* in the case of existing customers because the decision is based on their past buying and paying behavior (in the case of new customers *application scoring* takes place). The method used to derive rules for the decision can be a decision tree or a scoring scheme. Obviously, the relating data are generated and updated when a purchase or a payment takes place. But it is also important to update data when a customer with a balance due does not make a payment. Table 2 shows some of the data used in the example. The variable *number of successive months overdue* is a variable which value is updated monthly even if a customer has no transactions with the company in a given month.

Table 2: Customer data of a direct order company

CHAR005	Average of monthly balance
CHAR007	Balance of the actual month
CHAR008	Percent of used debt limit
CHAR013	Number of successive months overdue
CHAR022	Longest time of overdue in the last 12 month
...	...
CHAR516	Average of all payments
...	...
CHAR553	Percent of returns and refusals
CHAR555	Number of months since last return or refusal
...	...

6. SUMMARY

Table 3 gives an overview of the described data types with respect to some attributes. The values of the attributes should be understood as typical values.

Data type	Time of generation with respect to transactional data	Transaction dependent?	Ownership (ass.: company is owner of transaction data)
Circumstantial	at the same time	yes	others
Structural	before or after	no	company
Qualification	after	yes	company
Non-transactional tracking data	at any time	no	company or others

Table 3: Additional Data Types for Data Mining

Our findings are more specific with respect to types of data that can be used to enrich transactional and master data than general knowledge discovery process models but they are still not “cooking recipes” for data collection for data mining. Each case requires its own business analysis. Our findings may help to make management more aware of the data needs in data mining while data miners may find them helpful to conceptualize their thoughts.

REFERENCES

- [1] Chapman, P. et al., “CRISP-DM 1.0 - Step-by-step Data Mining Guide”, CRISP-DM consortium, 2000.
- [2] Fayyad, U., G. Piatetsky-Shapiro and P. Smyth, “From Data Mining to Knowledge Discovery in Databases“, AI Magazine, pp37-51, Fall 1996.
- [3] Fayyad, U and R. Uthurasamy, “Evolving Data Mining into Solutions for Insights”, Communications of ACM, Vol. 45, No. 8, pp28-31, 2002.
- [4] Hirji, K. K., “Exploring Data Mining Implementation“, Communications of the ACM, Vol. 44, No. 7, pp87-93, 2001.
- [5] Langner, R., Alpar, P. and Pfühl, M., “Ein Vergleich ausgewählter Klassifikationsverfahren im Kontext von Finanzdienstleistungen“, in „Wirtschaftsinformatik 2003“, Vol. II, Eds. Uhr, W., Esswein, W., Schoop, E. pp495-517, 2003.
- [6] Rapoport, R. N., “Three Dilemmas in Action Research“, Human Relations, Vol. 23, No. 4, pp499-513, 1970.
- [7] Wassermann, M., “Mining Data“, Federal Reserve Bank of Boston, Vol. 10, No. 3, Quarter 3 2000.
- [8] Wittmann, T., P. Kischka, M. Hunscher and J. Ruhland, “Data Mining - Entwicklung und Einsatz robuster Verfahren für betriebswirtschaftliche Anwendungen“, Frankfurt am Main, 2000.

A User-Oriented Contents Recommendation System in Peer-to-Peer Architecture

Hyeakyeong Kim, Jaekyeong Kim

School of Business Administration, Kyung Hee University, Seoul, 130-701, Korea
nell10@korea.com, jaek@khu.ac.kr

ABSTRACT

Recommender system is a popular technique for reducing information overload and finding digital contents that is most valuable to users. However, most recommender systems are based on a centralized client-server architecture in which servers and clients represents contents providers and users respectively. The existing recommender systems depend on contents providers and give a number of disadvantages to users. Therefore, we propose a recommender system based on a distributed P2P architecture that has originated with user-oriented principle rather than business itself. The proposed system consists of fully functioning personal recommender agents that automatically select neighbors and recommend contents. The agents learn user preference from users' content usage without requiring users' explicit ratings. We believe that the suggested P2P based recommender system should provide the users with more qualified recommendations, while it reduces the effort and time of users.

Keywords: Recommendation, Peer-to-Peer, Collaborative Filtering, Content-Based Image Retrieval

1. INTRODUCTION

According to a recent report, 93% of information produced worldwide is in digital form and the unique data added each year exceeds one exabyte, and more than 513 million people around the world are now connected to the global information resource [11]. However, all of those people have problems to search for digital contents they are most interested in. This trend calls for equally recommender systems with scalable searching capability. Recommender systems have been proved to be one of the most successful techniques to help people find contents that are most valuable to them in research and practice.

But, most existing Recommender systems are based on centralized client-server architecture in which servers and clients represent service providers and users respectively. The existing recommender systems depend on content providers and give a number of disadvantages to users. Centralized recommender systems collect users' sensitive information at one server. This makes recommender systems a serious risk for violating the privacy of users. Furthermore, from the users' perspective, the information about their preference is fragmented across many service providers reducing the quality of recommendations to them. Besides, each service provider requires users to indicate their specific preferences. So users must provide redundant information to obtain suitable recommendations from each service provider.

One solution to the problems with centralized recommender system is to distributed peer-to-peer (P2P) architecture, recent alternative to the dominating client-server architecture. Users in P2P architecture works as servers and clients simultaneously.

In this paper, we propose a collaborative filtering-based image content recommender system in distributed P2P architecture as a solution to the problems with centralized recommender system. Our system consists of users connected by personal recommender agents which reside on each user's computer. All processing to recommend contents is done locally by the agent. It allows users to maintain a fraction of the information about other users only, not all users' information, and generates recommendations themselves. Our proposed recommender agent selects more similar neighbors dynamically by learning preference, what image the user want, from the user's content usage without requiring user's explicit ratings. The system combines the two most popular information filtering techniques: Collaborative Filtering and Content-based Image Retrieval. This paper describes the reasons for and by which these two techniques were combined to recommend images in P2P architecture.

2. RELATED WORK

2.1 CF-based Recommender Systems

The recommender system is one of possible solutions to searching for individually preferred images from amount of images. A recommender system is defined as a system that assists customers in finding the items they would like to purchase. One of the most successful recommendation techniques is Collaborative Filtering (CF) [1,3,12,13], which has been widely used in a number of different applications.

Collaborative filtering is an information filtering technique that depends on human beings' evaluations of items. It is an attempt to automate the "word of mouth" recommendations. It identifies customers whose tastes are similar to those of a given customer and it

recommends items those customers have liked in the past. In general, CF-based recommender systems make recommendations according to the following steps [3,12]: (1) A customer provides the system with preference ratings on items that may be used to build a customer profile. (2) The system applies statistical or machine learning techniques to find a set of customers, known as neighbors, who had in the past exhibited similar behaviors. A neighborhood is formed based on the degree of similarity between a target customer and other customers. (3) Once a neighborhood is formed for a target customer, the system generates a set of items that the target customer is most likely to purchase by analyzing the items in which neighbors have shown an interest.

Despite their success, none of the previously proposed approaches can be an adequate solution for image recommendations in distributed P2P architecture. In this paper, Content-based Image Retrieval (CBIR) [5,6,10], the most widely used image retrieval technique, was used as a technique for content-based similarity search and was combined with CF.

2.2 Content-Based Image Retrieval

Content-based image retrieval represents an image as a point in multi-dimensional feature space and performs similarity-based retrieval using its visual features such as color, texture and shape. In CBIR, the customer describes visual characteristics of desired images using a query that is a set of example images. A query is internally represented as multiple points (i.e. query points) that have visual characteristics of example images. In general, CBIR systems retrieve images according to the following steps [15]: (1) A customer presents a query to the system via selection or sketching of images as a request for desired images. (2) The system searches for images similar to the query. The similarity between an image in the database and a query is calculated using the distance between corresponding points in the feature space. (3) The images with the highest degree of similarity are retrieved and recommended to the customer.

In spite of the virtues of CBIR in retrieving images similar to a query, CBIR rarely brings a customer to the desired images immediately. The reason for this is that any combination of example images may not precisely represent the images that a customer desires. For a system to handle this gap properly, it needs the ability to learn about what image the customer really wants through iterative interactions. The customer's current preference on the presented images needs to be fed back so that CBIR can learn from this preference to retrieve, in the next iteration, images more similar to the one customer really wants. This learning process, the *preference feedback*, is an essential mechanism for a faster search of desired images. We will refer to a set of preferred images as a *saved image set*. The images in

the saved image set are used for query refinement for the purpose of learning customer's current preference.

2.3 P2P (Peer-to-Peer) System

With the pervasive deployment of computers, P2P is increasingly receiving attention in research and practice. P2P system is direct communication or collaboration between two or more agents, such as personal computers or devices that bypass a centralized computer server. Current P2P applications can be classified into one of the following three categories [4]: Content File sharing, Distributed processing, Instant messaging

As the main usage of P2P systems are to share content files among a group of computers called peers in a distributed way, by direct exchange between peers. The using contents in P2P architecture differs markedly from the client-server architecture. A client-server based system depends on a single server storing information and distributing it to clients. The information repository remains essentially static, centralized at the server, and subject only to updates by the provider. Peers assume a passive role in that they receive, but do not contribute, information. A P2P system, on the other hand, considers all peers equal in their capacity for sharing information with other users. Each peer makes an information repository available for distribution, which, combined with anyone's ability to join the network, each peer can make information available for distribution and can establish direct connections with any other peers to download information. Instead of looking at what is available in a centralized repository, a peer seeking information from a P2P system searches across scattered collections stored at numerous peers, all of which appear to be a single repository with a single index. P2P system can use protocols that make it easier for each peer participate and share information, but the trade-off can be decreased quality of service.

Any P2P system doesn't provide recommendations to peers so that the peers have problems to search for contents that are most valuable to peers. To add value to P2P system, it needs to provide efficient search to peers to reduce information overload. We believe that P2P system with recommender functions make it possible. Therefore we propose contents recommender system in P2P architecture[2,8,9].

3. PROPOSED RECOMMENDER SYSTEM

Our system is designed to deal with the problems we face in recommending image files in a decentralized domain, P2P architecture. All processing to support a peer in finding desired images by generating personalized recommendations of image is done locally by a personal agent resides on each peer's computer without central server.

The system consists of interconnected peers of agents are implemented to be fully functioning recommenders. The agent keeps watch its peer (called the *host peer*)

and learn what images the host peer find relevant. That knowledge is then used to find both other relevant images and similar peers (called the *neighbor set*). The agent exchanges recommendations with the agents of its neighbor set.

The personal agent consists of two modules: CF agent and CBIR agent. Two components collaborate each other to decide whom should receive a recommendation, whom to keep as a neighbor, and what contents relevance to the host peer. Figure 1 shows proposed recommender agent with overall procedure.

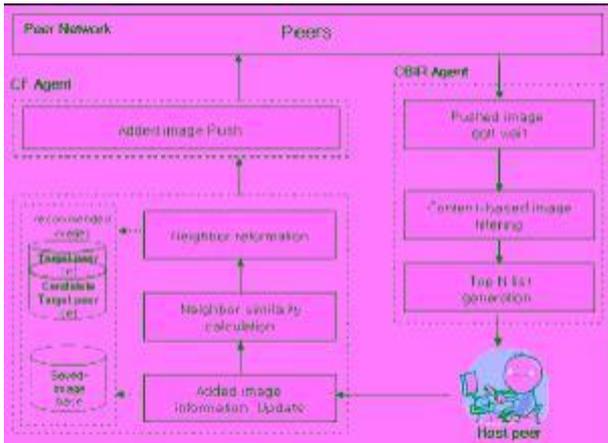


Figure 1. Personal Recommender Agent

The agent dynamically selects which peers to include its neighbor set. When a peer is discovered to frequently provide good recommendations, the peer attempts to move closer to the host peer. Thus an agent might keep better constituent neighbors and recommendations should be reached to more related peers. This mechanism provides a wide peer network with efficient image sharing being spread faster with less communications.

3.1 Peer Model

In P2P architecture, a personal recommender agent builds a peer model respectively to generate personalized recommendations for a host peer. The peer model should include information about a host peer’s preference for image and its neighbors to exchange recommendations with other agents.

Image representation: To analyze peers’ preference for images, we use the HSV (i.e. hue, saturation, and value of color) based color moment over other choices of features such as shape or texture, because color moment is the most generally used feature and HSV represents human color perception more uniformly than others [9]. For all pixels in images, we translated the values of three-color channels (i.e. RGB: red, green, and blue) into HSV values. Then, the mean, standard deviation and skewness for HSV values were calculated to represent images as vectors in a nine dimensional feature space.

Peer Profile: A peer profile consists of two parts: saved image set and neighbor image set. A set of preferred images is called the saved image set, $Q = (q_1, q_2, \dots, q_j)$, denotes all images that host peer has saved, which is used to refine the query for the purpose of learning host peer’s current preference. The neighbor image set, $P_{ij} = (p_{i1}, p_{i2}, \dots, p_{ik})$, denotes image j recommended by neighbor i which is used to calculate similarity of the neighbor i . The neighbor image set includes images a host peer has saved only.

Saved image set includes information about peer’s preferences on images, while neighbor image set includes information about similarity of each neighbor according to their recommendations. The peer profile is constantly updated with newly obtained relevance images and similarity of its neighbors to dynamically reflect peer’s most recent preference. This is significantly different from the user profiles used in traditional server based recommender systems [11].

3.2 Neighbor Definition

The agent of host peer h has a neighbor set N_h consists of peers who have recommended relevance images to host peer h . Each Neighbor has a $s(n_k, h)$ respectively. Here, $s(n_k, h)$ denotes the similarity between a neighbor n_k and host peer h . As equation (1), a similarity accumulates the reciprocal of the $Dist(p, Q)$ calculated by CBIR agent is discussed in section 4.2 to reflect both the similarity and the success times of recommendations.

$$s_{i+1}(n_k, h) = s_i(n_k, h) + Dist(p, Q^h)^{-1} \quad (1)$$

The neighbors of host peer h divided into two sets: the target peer set (N_h^t) and the candidate target peer set (N_h^c).

$$N_h = N_h^t \cup N_h^c \quad (2)$$

Using the $s(n_k, h)$, the agent generates a set of target peers, $N_h^t = \{n_1, n_2, \dots, n_k\}$, such that $s(n_1, h)$ is maximum, $s(n_2, h)$ is the next maximum, and so on where k is a limited number of target peers. The agent of a host peer forwards recommendations to its target peers only. Besides target peers, the rest of neighbors is included to candidate target peer set (N_h^c). But, neighbors are dynamically exchanged: adding new peers and moving between two sets to reflect host peer’s current preference.

4. RECOMMENDER AGENT

The personal agent consists of two modules: CF agent, CBIR agent. In this section, we describe the roles of two agents respectively.

4.1 CF Agent

CF agent is implemented with a learning algorithm to model a host peer and with a set of neighbor peers. CF agent makes it possible to find more relevance neighbors. The agent updates the peer profile using feedback information (*saved or not*). The updated profile is used to choose whom to keep as neighbors and target peers but also refine the query. The refined query is passed to the CBIR agent to predict the relevance of images for current preference of the host peer. CF agent can only generate recommendations of its host peer, and forward other peers' recommendations to its target peers.

Initialization To initialize a new peer, agents can use a different algorithm. This is implemented by initializing the peer with neighbors taken from an artificially created. To determine initial neighbor set for new host peer p , the agent calculates the similarity between new host peer p and other peers using all saved image set as follows:

The agent determine initial neighbors for a peer p , $B = \{b_1, b_2, \dots, b_m\}$ such that $p \notin B$ and $sim(p, b_1)$ is maximum, $sim(p, b_2)$ is the next maximum, and so on. Here, $sim(p, b)$ denotes the similarity between two peers p and b , and is calculated using equation [3][4].

$$Dist(P, Q) = \sqrt{\frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} dist^2(p_i, q_j)}{n_1 n_2}} \quad (3)$$

$$dist(p_i, q_j) = \sqrt{\sum_{s=1}^S (p_{is} - q_{js})^2} \quad (4)$$

Where n_1 and n_2 are the number of saved images of the peer b and peer p respectively, and p_i and q_j are the i th and j th image, and $dist(p_i, q_j)$ is a distance function between an image p_i and q_j . And in equation (4), s is the number of dimensions of the feature space and p_{is} and q_{js} are coordinates of an image p and q on the s th dimension respectively. It has characteristics of treating neighbors with similar saved images using the shortest distance.

The set of similar peers to new peer p , $B = \{b_1, b_2, \dots, b_m\}$, will be the agent's initial neighbor set. Additional neighbors will be found by the exchange of recommendations.

Recommendation Generation An agent generates recommendations to each other not on request but whenever a host peer saves a new image. However, each agent can only collaborate directly with its target peers, and hence to reach agent beyond the target peers they have to forward recommendations for each other. Figure 2 illustrates it.

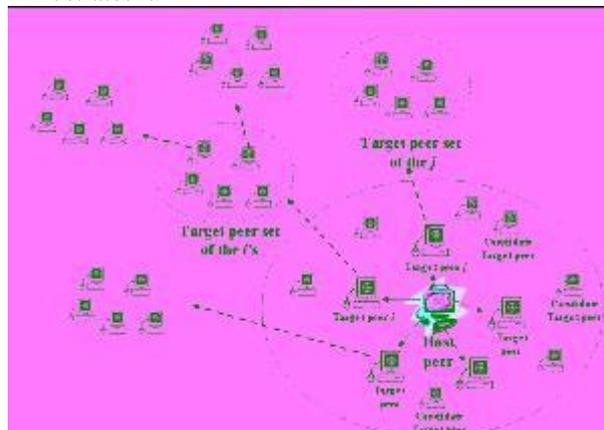


Figure 2. Communications of an agent with its target peer set

Neighbor Reformation The recommender agent dynamically selects which users to be included in neighbors. When a host peer receives a suitable recommendation, the recommending peer might be added to neighbor set. When an agent receives a forwarded recommendation from other peer i , the recommending peer i might be added to the candidate target peer set and discards an old one j with the lowest similarity value as equation (5).

$$\begin{aligned} & \text{if } i \notin N \\ & \text{then } N_c \leftarrow N_c - \{j\} + \{i\} \end{aligned} \quad (5)$$

Moreover, if a candidate target peer recommends relevant images continuously, it becomes a target peer. If the candidate target peer is more similar to the host peers than the target peer, the candidate target peer will replace the target peer and vice versa in each neighbor set.

$$\begin{aligned} & \text{if } \text{arc max}_{n_{c_j} \in N_c} (sim(h, n_{c_j})) > \text{arc min}_{n_{t_i} \in N_t} (sim(h, n_{t_i})) \\ & \text{then } N_c \leftarrow N_c - \{n_{c_j}\} + \{n_{t_i}\} \ \& \ N_t \leftarrow N_t - \{n_{t_i}\} + \{n_{c_j}\} \end{aligned} \quad (6)$$

According to this mechanism, an agent finds more similar neighbors resulting to be given relevant images. Thus an agent might keep better constituent neighbor members and recommendations should be reached to more similar users.

4.2 CBIR Agent

A recommender agent can receive from other recommender agent. All received images are put in a queue except the same as saved images. The CBIR

agent uses all saved images of the host peer as multiple query points, and retrieves images based on distance between the multiple query points and images in the queue. For all images in the queue, this agent calculates the distances from the multiple query points and generates a list of k nearest images (k -NNs) as recommendations. This agent retrieves k images entirely based on visual features of query that represents the peer's current preference.

The calculation to retrieves k images is as follows: CBIR agent uses the all saved images as the query Q , and continuously refines the query by adding the query points in Q with the newly saved images. Since a query is allowed to have multiple query points, the distance function between an image x and a query Q should aggregate multiple distance components from each query points to the image. We use the following aggregate distance function:

$$Dist(x, Q) = \sqrt{\frac{g}{\sum_{j=1}^g 1/dist^2(x, q_j)}} \quad (7)$$

where g is the number of query points in a query Q , q_j is the j th query point of Q , and $dist(x, q_j)$ is a distance function between an image x and a query point q_j . We derived the Equation (7) from the FALCON's formula [14]. It has characteristics of treating an image with the shortest distance component to any one of query points as the image with the shortest aggregate distance.

Just as this system continuously refines the query in newly added images, it also updates the Equation (7) to reflect the peer's current preference. For this purpose, we define the $dist(x, q_j)$ in Equation (7) as:

$$dist(x, q_j) = \sqrt{\sum_{l=1}^L w_l (x_l - q_{jl})^2} \quad (8)$$

where L is the number of dimensions of feature space, w_l is a weight of the l th dimension in the feature space, and x_l and q_{jl} are coordinates of an image x and a query point q_j on the l th dimension, respectively. w_l in Equation (8) is substituted by $1/s_l$ at the end of every connection of the peer, where s_l is a standard deviation of coordinates of l th dimension of images. Note that s_l is calculated using all images in the relevant set accumulated. This distance function update is to better reflect a peer's current preference by allowing different weights by dimension and putting more emphasis on the features with smaller variance. Note that equal weights are to be used for all dimensions when there is only one query point in Q .

Based on the above definitions and discussions, CBIR agent generates Top k recommendation list for the host

peer, $X=\{x_1, x_2 \dots x_k\}$ such that $Dist(x_1, Q)$ is the minimum, $Dist(x_2, Q)$ is the next minimum, and so on.

The retrieved k images are presented when the peer connects the network. And the peer skims through the list to see if there are any images of interest. Then, the peer may save desired images on the peer's computer. This relevance judgment on the presented images passes to CF agent to learn the peer's current preference.

5. EVALUATION

In order to performance evaluation the proposed image recommender system we implement a simulator with variable parameters: size of the neighbor set, exploration propensity of the user for simulated 30 days. The components of the simulator are some parts of the recommender system. One of the simulated parts is images including their representation as vectors in a nine dimensional feature space for HSV values. For the simulations, the 1000 images that Korea Telecom Freetel (KTF), a leading Korean CDMA carrier, is currently offering were used. Other parts consist of 1000 peers modeled as having distinctive preference for some images. And they have initial neighbor set to exchange recommendations. Last, the recommender agents work as described in section 4.

Proposed recommender system is compared with centralized recommender system. In our system, each peer has a unique recommender agent with a limited neighbor set, while in the compared system, a single agent manages all users for recommendations. For performance evaluation of the system, we use metrics most widely used for recommender systems, *precision* and *recall* defined as follows:

$$Precision = \frac{\text{recommended relevant images}}{\text{all recommended images}}$$

$$Recall = \frac{\text{recommended relevant images}}{\text{total relevant images}}$$

The quality of recommendations is known to vary by the size of the neighbor set, the number of TTL. The metrics for each individual peer were computed and the average value for use as the metric was calculated. We watch how performance of the system changes for every simulation day.

For analysis of experiment results, two statistical tests are conducted. One is the t test for comparison of the average performance of our system with the centralized system. The other is the two-way ANOVA test with repetition for assessment of the effects of two factors, size of the neighbor set and n , on t .

6. CONCLUSION

We propose a recommender system based on a distributed P2P architecture that has originated with

user-oriented principle rather than business itself. This could extend research area of the recommender systems to distributed architecture.

We develop an adequate solution for image recommendations in distributed P2P architecture, combines two techniques: collaborative filtering and content-based image retrieval. The system supports peers of the P2P system in finding a desired image by generating personalized recommendations of images. We implemented fully functioning personal recommender agent that automatically selects neighbors and recommends contents by learning user preference from user's content usage without requiring user's explicit ratings. We believe that the suggested P2P based recommender system should provide the users with more qualified recommendations, while it reduces the effort and time of users.

The proposed system has flexible algorithm to share any contents to can be analyzed. Therefore our future work includes varying contents, such as music and text, and implementing other techniques besides CBIR for image.

REFERENCES

- [1] Balabanovic, M., Y. Shoham, "Fab: Content-based Collaborative Recommendation", *Communication of the ACM*, Vol. 40, No. 3, pp 66-72, 1997.
- [2] Canny, J., "Collaborative filtering with privacy", *In Proceedings of the IEEE Symposium on Research in Security and Privacy*, pp 45— 57, 2002.
- [3] Cho, Y.H., J.K. Kim, S.H. Kim, "A Personalized Recommender System Based on Web Usage Mining and Decision Tree Induction", *Expert Systems with Applications*, Vol. 23, No. 3, pp 329-342, 2002.
- [4] Damiani, E., S. De Capitani di Vimercati, S. Paraboschi, P. Samarati, F. Violante, "A reputation-based approach for choosing reliable resources in peer-to-peer networks", *In Proc. of the 9th ACM Conference on Computer and Communications Security*, pp 17-21, November 2002.
- [5] Kim, C.Y., J.K. Lee, Y.H. Cho, D.H. Kim, "VISCORS: a Visual Contents Recommender System on the Mobile Web", *IEEE Intelligent Systems, Special issue on Mining the Web for Actionable Knowledge*, Accepted for publication, 2004.
- [6] Kim D.H., C.W. Chung, "Qcluster: Relevance Feedback Using Adaptive Clustering for Content-Based Image Retrieval", *Proc. ACM SIGMOD International Conference*, pp 599-610, 2003.
- [7] Melville, P., R.J. Mooney, R. Nagarajan, "Content-boosted collaborative filtering for improved recommendations", *Proc. the 18th National Conference on Artificial Intelligence*, pp187-192, 2002.
- [8] Olsson, T., "Bootstrapping and Decentralizing Recommender Systems", *PhD thesis, Uppsala University, Uppsala, Sweden*, 2003.
- [9] Peng, H., X. Bo, Y. Fan, S. Ruimin, "A scalable P2P recommender system based on distributed collaborative filtering", *Expert Systems with Applications*, Vol. 27, pp 203–210, 2004.
- [10] Porkaew K., K. Chakrabarti, S. Mehrotra, "Query Refinement for Multimedia Similarity Retrieval in MARS", *Proc. 7th ACM Multimedia Conference*, pp 235-238, 1999.
- [11] Prete, C.D., J.T. McArthur, R.L. Villars, I.L. Nathan Redmond, D. Reinsel, "Industry developments and models, Disruptive Innovation in Enterprise Computing: storage", *IDC*, 2003.
- [12] Sarwar, B., G. Karypis, J.A. Konstan, J. Riedl, "Analysis of recommendation algorithms for e-commerce", *Proceedings of the ACM Conference on Electronic Commerce*, pp 158-167, 2000.
- [13] Schafer, J.B., J.A. Konstan, J. Riedl, "E-commerce recommendation applications", *Data Mining and Knowledge Discovery*, Vol. 5, No. 1-2, pp 115-153, 2001.
- [14] Wu, L., C. Faloutsos, K. Sycara, T. Payne, "FALCON: Feedback Adaptive Loop for Content-Based Retrieval", *Proc. 26th VLDB Conference*, pp 297-306, 2002.
- [15] Zhou, X.S., T.S. Huang, "Relevance feedback for image retrieval: a comprehensive review," *ACM Multimedia Systems Journal*, Vol. 8, No. 6, pp 536-544, 2003.

An Architecture of FoaF-based Peer 2 Peer Knowledge Management System

Xuan Zhou¹, Jian Zhou¹, Qing Li¹, Yuliu Chen¹, Lars Ludwig²

¹ Department of Automation, Tsinghua University, Beijing, 100084, China

² DERI, Nation National University of Ireland, Galway, Ireland

xuan.zhou@deri.org

ABSTRACT

A number of researches have been focused on improving the productivity of knowledge workers. Effective communication is determinative, and one important problem has been to find the right person for collaboration. The Semantic Web efforts improve it by better wide-range knowledge management. Through extending the FoaF (Friend of a Friend) vocabulary, knowledge workers can share their machine-readable knowledge world widely, so that their PCs can facilitate effective collaboration efficiently. This paper presents an approach to support networked personal knowledge management. An example of extended FoaF is also described

Keywords: FoaF, Social Network, Peer 2 Peer, Knowledge Management

1. INTRODUCTION

Efforts in improving the productivity of knowledge workers are on the cutting edge of technology development. Recent research has proved that social network is important for the successful generation, acquisition, diffusion, and exploitation of knowledge, which is the keys to strong performance in the Knowledge Based Economy (KBE). Various efforts were devoted to discover and release the “strength of weak ties”[1]. By adding metadata management capacity, Peer 2 Peer networks are in the early stage in evolving its application from ‘file sharing’ to ‘collaboration infrastructure’[2]. In the mean time, FoaF is being widely used to connect people over the Internet. However, by only sharing the name, address etc. is inadequate for effective and efficient collaboration among peers. Furthermore, all these initiatives are facing challenge due to the emergence of Semantic Web (i.e. the next generation web) [3]. Therefore, it is necessary to investigate how semantic technology can facilitate Peer 2 Peer application and personal knowledge management, and improve the productivity of knowledge workers consequently.

This paper presents an application architecture with the purpose of manage personal knowledge management in social network environment. It manages the knowledge generated by user, structures collaboration-relevant information with an extended FoaF vocabulary, and publishes them into shared space. In addition, agents are used to make some processes automatic.

The structure of the paper is as follows. In section 2, theory foundation such as social network, knowledge workers, and FoaF are introduced and discussed. Afterward, a FoaF-based approach for networked personal knowledge management is presented in section

3. Finally section 4 concludes the paper by discussing the applicability of this approach and potential further research.

2. SOCIAL NETWORKING AND FOAF

This section focuses on the human side of knowledge application and innovation management. At first, the importance of social network is briefly introduced and how FoaF can support collaboration is discussed at last.

2.1 Social Networks

There have already been several organization structures so far (see Fig 1). Despite that functional organization is still the basic structure of enterprise, a number of organization structures, such as market-oriented, team-based, cop-based, appeared as a result of seeking comparative and competitive advantage.

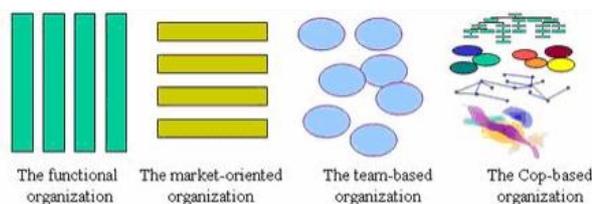


Fig. 1. Structures of organization (source: COP Consortium 2000)

Even in a traditional function-oriented organization (e.g. a corporation), there are various informal organizational channels, which is important for the internalization of organizational knowledge and externalization of individual knowledge. Fig 2 illustrates the difference between formal and informal organizational structure.

¹ Sponsored by China 863 Program No : 2001AA415340 and China National Natural Science Fund No : 60474060

In other words, there are no further information about the friend's value, goal, current plan etc. Furthermore, the exiting class and property in FoaF vocabulary can not carry the information for collaboration. An important feature of FoaF is that its vocabulary (classes and properties) can be extended according to requirement. For example, some researchers in DERI (www.deri.org) are extending the FoaF vocabulary for a semantic web portal [9], and there is also an effort in defining often changing variables in FoaF [10]. However, little attention has been conducted on FoaF extension for improving the productivity of knowledge workers. Therefore this study proposes a corresponding approach, and an example of FoaF extension is presented in section 3.

3. APPLICATION FRAMEWORK BASED ON EXTENDED FOAF

Based on previous analysis, it is possible to extend FoaF vocabulary and build an application in pursuit of the productivity of knowledge worker community. In this section, the application background and methodological framework are introduced, and some design specifications for personal information management in a community context is also presented.

3.1 Application Background

There are lots of applications dedicated to managing personal knowledge, and most of them are document based. The Internet/intranet is also a carrier of knowledge, which became more and more popular due to its convenience. As for the system, it is generally accepted that lifecycle support feature can largely affect the effectiveness. Fig. 4 illustrates one paradigm for individual knowledge management.

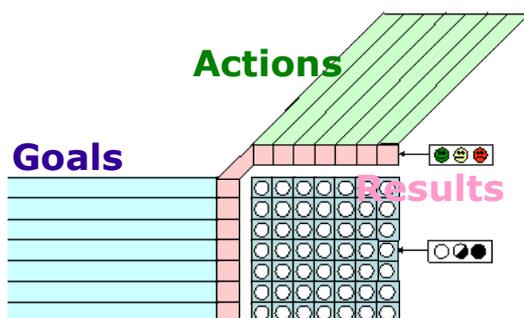


Fig. 4. Personal knowledge management matrix

Each goal, action, and their relationship are measured by qualitative indicators (see the following step 4). Thus, Goals, Actions, and their relations form a 'matrix', which is an excellent channel to monitor in order to achieve individual efficiency and effectiveness. Furthermore, sharing the general description of goals and actions will facilitate knowledge reuse, improve collaboration consequently. There are four steps in utilization of this matrix:

Step 1. Define personal Goals: Identify and populate the most important Goals. Goals are defined as the objectives of effort. They are also known as strategies or visions (e.g.).

Step 2. Define personal Actions: Identify and populate the most important Actions. Actions are defined as expenditure of effort and are also known as projects or tasks (e.g.).

Step 3. Link Goals with Actions: Indicate the strength of the relationship between each action and goal. An empty box indicates no relationship; a half full box indicates a partial relationship; and a full box indicates a strong relationship (i.e. completing the action will have a high impact on attaining that particular goal).

Step 4. Indicate the status of Goals and Actions: Continuously monitor and report the status of goals and actions. In this case, red indicate that urgent attention is required, yellow indicates that more work is needed and green indicates that the goal or action requires no immediate attention.

After finish all the steps, we will get a picture as Fig 5 illustrates.

Actions [New]	Goals [New]			Help
	Goal1	Goal3	Goal2	
Action1	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	●
Action2	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	●
Action4	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	●
Action3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	●
Results	●	●	●	

Fig. 5. Example of personal knowledge management modules

It should be noted that this methodology is based on the concepts which any knowledge workers are sharing. However, in order to be proficient in it, one must practise projecting his ideas into the matrix items. Moreover, if any action does not have relationship with any goal, or if any goal does not have relationship with any action, their necessity needs to be reconsidered. Thus an iterative process is necessary for accuracy.

3.2 Information Exchange Framework

A personal knowledge management paradigm, described in section 3.1, is established by three major objects. However, not every piece of information from them is good for sharing. In fact, information about the 'result' is designed for individual feedback control. Therefore it is necessary to make a distinction between the two kinds of information. In addition, the data format will be likely different since the consumer of them are inhomogeneous. Furthermore, the data will accumulate with time going on, and publish publicly the past data will not make any sense. Fig 6 illustrate a model for information exchange.

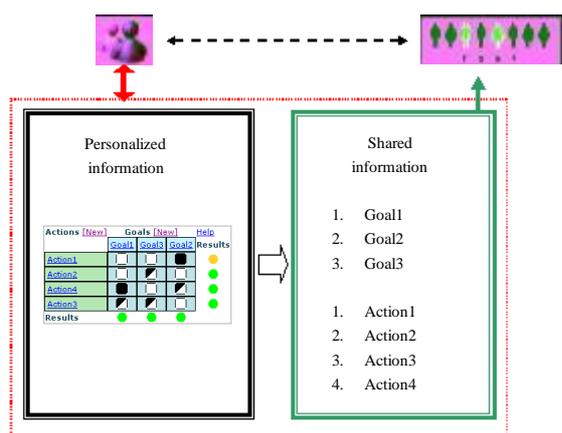


Fig. 6. Information for personal knowledge management and exchange model

From Fig 6, we can see, information is divided into two categories: Personalized information and Shared information:

-*Personalized information*: operated with traditional user interface; can be located in a PC or a repository on a server; and contains all the data, including past data.

-*Shared information*: stored in a web server which is accessible to the public; in standard format (in this case, it is an extended FoaF); and only contains the current published data.

Knowledge workers will change their personalized information occasionally, while shared information, as a subset of Personalized information, will impel collaboration among knowledge workers. In this way, one can find a friend with interesting goal or similar action through extended FoaF. At the mean time, everyone can update their personal information easily with their traditional tools (e.g. Planplus from franklincovey). Fig 7 illustrates this scenario.

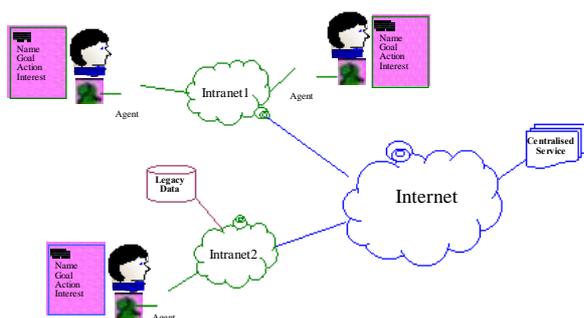


Fig. 7. An framework of networked personal knowledge management system

In this design, data are structured with extended FoaF vocabulary. Thus we can share straightly information based on FoaF files, distributed across web servers. However, the data are stored in plain text with human language, which cannot be directly understood by machines. In order to make machine manage more tasks for users, agents are necessary for interaction on behalf of ‘friends’(see Fig7). Therefore, the basic roles of agent

should consist of: (1)Publishing- provide information to user’s FoaF file; (2)Searching- search in the FoaF files across the web; (3)Reasoning- analysis the language to discover the relationship between different pieces of data; (4)Contact- send and receive message to and from other people.

3.3 Extension of FoaF

The individual information about ‘goal’and ‘action’is similar to ‘interest’and ‘plan’, which are two developing concept included in FoaF vocabulary. However, the purpose of them are not the same, ‘interest’and ‘plan’ are more about personal life, which could be ‘water surfing’, ‘go to Valencia for vacation’etc. In other words, they are not mission critical, and it is necessary to extend FoaF vocabulary to facilitate collaboration by sharing individual goals and actions(see Fig 8).

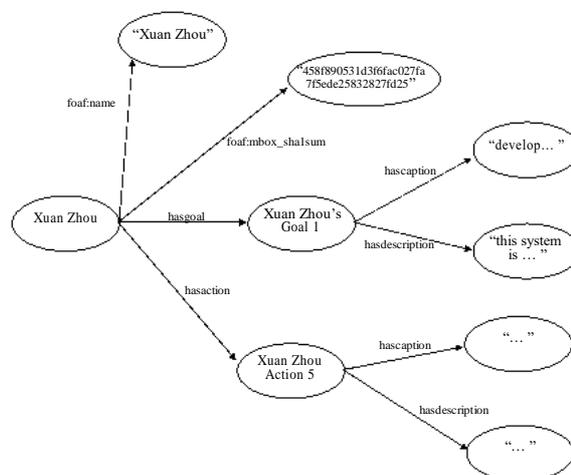


Fig. 8. RDF description of FoaF extension

There are two classes created, i.e. goal and action (in Fig8, they are two instance). They are linked to the person with two properties: ‘hasgoal’and ‘hasaction’. And each goal has two properties: ‘hascaption’and ‘hasdescription’. The content of goal or action are strings stored in these two properties. Table1 is an example of vocabulary extension in FoaF.

With extended vocabulary, we can now build the machine-readable FoaF file for collaboration. Table 2 is an example of the content of an extended FoaF file.

In this incomplete example, there are one instance of goal and one instance of action. The content of them are stored in two properties respectively (the description is omitted for concision). This is a first step for knowledge worker to collaborate over Internet. However, more artificial intelligence can be added to the initiative by define the connection between goals, actions. For example, friend1:goal3 is associated with friend8:aciton2. In addition, it is also reasonable to define several relationships for the connection.

Table 1. Foaf vocabulary extension

<p>Class: foaf:Goal <i>Goal</i> –An objective of this person. The foaf:Goal class represents an objective of a foaf:Agent.</p> <p>Class: foaf:Action <i>Action</i> –An objective of this person. The foaf:Action class represents an action of a foaf:Agent.</p> <p>Property: foaf:hasgoal <i>define</i> –something which is defined by this person. Status: Testing Domain: foaf:Person Range: foaf:goal The foaf:hasgoal property represents an objective of a foaf:Agent.</p> <p>Property: foaf:hasaction <i>define</i> –something which is defined by this person. Status: Testing Domain: foaf:Person Range: foaf:action The foaf:hasaction property represents an objective of a foaf:Agent.</p> <p>Property: foaf:hascaption <i>define</i> –something which is defined by this person. Status: Testing Domain: foaf:goal Range: foaf:goal The foaf:hasgoal property represents an objective of a foaf:Agent.</p>

Table 2. Content example of an extended Foaf file

<pre> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:foaf="http://xmlns.com/foaf/0.1/" <foaf:Person rdf:nodeID="me"> <foaf:name>Xuan Zhou </foaf:name> <foaf:mbox_sha1sum>458f890531d3f6fac027fa7f5ede258328 27fd25</foaf:mbox_sha1sum> <foaf:hasgoal> <foaf:goal> <foaf:hasname>p2p knowledge management</foaf:hasname> <foaf:hasdescription>... </foaf:hasdescription> </foaf:goal> </foaf:hasgoal> <foaf:hasaction> <foaf:action> <foaf:hasname> design the architecture </foaf:hasname> <foaf:hasdescription>... </foaf:hasdescription> </foaf:action> </foaf:hasaction> </foaf:Person> </rdf:RDF> </pre>

4. CONCLUSION

The productivity of knowledge workers are impacted by both technology and social aspects. The approach presented in this paper extended the Foaf vocabulary to support personal knowledge management in a networked environment. Knowledge workers will benefit from a dedicated front-end application for managing their personalized knowledge and an open interface (extended Foaf) for managing their shared knowledge.

It should be noted that this framework is mainly a conceptual and logic design, and can be easily used to setup a centralized service(e.g. within a corporation) or improve it. However, to implement it in totally distributed way(e.g. internet) is another pair of shoes, because several underground technologies are still in their developing phase. In addition, trust relationships between knowledge workers are another practical issue needs further discussion.

Notwithstanding its limitation, this study does suggest a framework, which provides a channel for knowledge workers to collaborate more productively over the Internet. It can improve the productivity of researchers in public section (e.g. universities) immediately after implemented and further research and development will make it more practical and effective.

ACKNOWLEDGEMENT

The authors would like to thank Knud Möller and Dieter Wolf for some important discussions about the issues addressed in this paper. The research was supported by a fellowship from Science Foundation Ireland.

REFERENCES

- [1] Granovetter, M., *The Strength of Weak Ties*. American Journal of Sociology, 1973(78): p. 1360~1380.
- [2] Decker, S. and M. Frank, *The Networked Semantic Desktop*. 2004.
- [3] Fensel, D., et al., *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. 2003: MIT Press.
- [4] Cross, R., A. Parker, and S.P. Borgatti, *A bird's-eye view: Using social network analysis to improve knowledge creation and sharing*. 2002, IBM Institute for Business Value.
- [5] Granovetter, M., and Patrick McGuire, *The Making of an Industry: Electricity in the United States*, in *The Law of Markets*, M. Callon, Editor. 1998, Blackwell: Oxford. p. 147-173.
- [6] Fensel, D., *Triple-based Computing*. 2004.
- [7] Dan Brickley, L.M., *FOAF Vocabulary Specification*. 2004.
- [8] Dumbill, E., *XML Watch: Finding friends with XML and RDF*. 2002.
- [9] Moller, K., L. Predoiu, and D. Bachlechner, *Portal Ontology*. 2004.
- [10] Gan, J.D., B.K. DeLong, and C. Schmidt, *A FOAF extension for defining often changing variables in FOAF*. 2004.

Distributed Load Balancing in Peer-to-Peer Computing

Shan Zhang, Zheng Qin

Software School, Tsinghua University, Beijing, 100084, China

s-zhang@mails.tsinghua.edu.cn, qingzh@tsinghua.edu.cn

ABSTRACT

In this paper, we address the load balancing problem in the context of peer-to-peer computing environments. The key challenge to employ peer-to-peer networks for distributed computing is to exploit the heterogeneous processing capability of the participating hosts as well as the diverse network conditions. The contribution of our work is twofold. First, we model the load balance problem as an optimization problem with the objective of minimizing the system response time. This modeling considers not only the current loading of hosts, but also the fluctuation of network delay, which completely captures the characteristics of the P2P systems. Second, we propose a gradient projection algorithm to solve the optimization problem, which is fully distributed and easy for implementation. Simulation results demonstrate that our scheme has satisfied performance in terms of convergence, response time and load distribution.

Keywords: peer-to-peer computing, load balancing, constrained optimization, gradient projection

1. INTRODUCTON

The Peer-to-Peer (P2P) systems are composed of a set of end hosts that work collaboratively for some common tasks, such as file sharing, on-demand streaming, collaborative computing etc.. They differ from the traditional distributed systems from several aspects. First, the participating hosts in the P2P systems are often distributed across the Internet and higher end-to-end delay is expected between logically closet neighbor nodes. Second, most of these nodes are normal PCs and have diverse processing speed, storage capacity and access bandwidth. These features make it challenging task to design the load balancing scheme to coordinate the distribution of items to be stored, data to be downloaded, or computations to be carried out to the participating hosts so that the heterogeneity of the hosts can be fully exploited.

Addressing this problem, several solutions have been proposed in literature [1]-[4]. These schemes are focused on the load balancing for structured P2P systems. Structured P2P systems employ a class of distributed hash tables (DHT) for item storage and retrieval. The main limitation of this kind of systems is that the distributed hash tables may not produce load balance as good as standard hash function, so there is a $\Theta(\log(N))$ imbalance factor in the number of items stored at a node [2]. In [2], Rao et. al. assume that the system is static, i.e., the membership and data items are stable over a timescale that the load balancing algorithm can be performed. They employ the concept of virtual server that represents a peer in DHT. Each physical node can host one or more virtual servers so that the storage and retrieval of data items are conducted at the virtual server level instead of the physical node

level. The key to achieve load balance is then to move virtual servers from overloaded physical nodes to lightly loaded physical nodes. This concept is extended to the dynamic P2P systems by Godfrey et. al. in [4], where data items can be continuously inserted and deleted, and nodes may change their states frequently. They propose to maintain the loading information of the peer nodes on a set of directory servers. These directory servers periodically schedule the reassignment of the virtual servers among the physical nodes that they are serving. However, these solutions do not consider the cost to move the virtual servers between the hosts. In [5], Zhu and Hu proposed a proximity-aware load balancing scheme to address this issue. The basic idea is to utilize the proximity information of the physical nodes, so that the virtual servers are only reassigned and transferred between physically close heavy nodes and light nodes so that the cost of performing reassignments is reduced.

The drawback of the virtual server based load balancing schemes is that running multiple virtual nodes may increase the bandwidth consumption in maintaining the network connectivity across neighbor nodes in the peer-to-peer network. In [6], Karger et. al. address this problem by arranging for each node to activate only one of its virtual nodes at any given time. Specifically, the node checks its inactive virtual nodes periodically. If it finds that the load distribution has changed, it may migrate to one of these virtual nodes. In this way, previous load is shifted to other nodes, and it will take responsibility for loading of the newly migrated virtual nodes.

All of the aforementioned schemes are based on the structured P2P systems where the load distribution is tightly controlled by the distribution hash functions. Although the virtual server concept is elegant and easy

to implement, the achieved load balancing is ad-hoc in that it relies on the reassignment or migration of data items after they have been assigned. The fineness depends on the space allocated for each virtual server. The tradeoff, however, is not easy to justify. Furthermore, the scheme adopted in [4] essentially resorts to the centralized solution, which makes the directory servers vulnerable to single-point-failure.

In this paper, we consider the load balancing problem in the context of unstructured P2P systems. Example of this kind of systems includes Gnutella, KaZaA, etc. A salient feature of these unstructured P2P systems is that the host can fully control where its data items can be placed, and the amount of the data items that should be assigned to these hosts. This feature is desirable for load balancing purpose since it allows for a fine-grained allocation of load across all participating hosts according to their loading. However, the results is not only limited to unstructured P2P system, it is also applicable for structured p2p system. For example, it can be used to find a finegrained reassignment of virtual servers for this kind of p2p systems.

The contribution of our work is twofold. First, we model the load balance problem as the minimum cost problem. This modeling considers not only the current loading state of the participating hosts, but also the fluctuation of network delay, which completely captures the characteristics of the P2P systems when they are employed for massive computing purpose. Second, we propose a gradient projection algorithm to solve the optimization problem, which is fully distributed and easy for implementation.

In the rest of this paper, we first present the system model and problem formulation in Section 2 and discuss the necessary conditions for the optimization problem. In Section 3, we then propose an algorithm based on the classic gradient projection methods. By identifying the difficulties to implement the algorithm in practice, we propose an approximation solution which is suited for practical implementation. We also discuss the inherit reasonability of our model for the p2p computing environments. Simulation results are presented in Section 4 and finally we conclude this paper in Section 5 and suggest some possible directions for future research.

2. SYSTEM MODEL AND PROBLEM FORMULATION

2.1 Problem Statement

We are given a peer-to-peer system and a set of N hosts. Each host $i \in N$ can choose a set of S_i hosts from these N hosts as its servers to carry out the computation for its raw data. These servers can be chosen randomly as in the unstructured peer-to-peer networks, or can be determined according to the distributed hash function as that of the structured peer-to-peer networks.

For each host i , we are given a scalar r_i referred to as the generating rate of the raw data of host

i to be computed. In the context of the peer-to-peer computing, r_i (measured in units/second) could be the data rate exporting from host i to its servers for computation for the final results. The objective of our problem is to divide r_i of host i among the set of S_i servers in a way that the resulting loading distribution across the overall peer-to-peer system minimizes a suitable cost function.

We denote x_{ik} as the data rate that is assigned by host i to its server $k \in S_i$. The collection of assignment x_{ik} must satisfy the following constraints

$$\sum_{k \in S_i} x_{ik} = r_i, \quad \forall i \in N \quad (1)$$

$$x_{ik} \geq 0, \quad \forall k \in S_i, i \in N \quad (2)$$

For peer-to-peer computing, the primary design objective is to minimize the response time for the tasks, which is the duration from the time that a host assigns the raw data to the server to the time that the final result is computed and retrieved. In peer-to-peer systems, the participating hosts have highly diverse processing speeds and are widely distributed across the Internet. As a result, the response time is composed two components, one is host dependent processing delay and the other is network dependent end-to-end delay.

For host-dependent delay, we adopt the response time model from [7] by treating the computing process as an M/M/1 system. Specifically, suppose a host k is chosen by a set of A_k hosts as their servers. Then the aggregated data rate to be carried by this host is

$$J_k = \sum_{i \in A_k} x_{ik} \quad (3)$$

Suppose the host k has a processing speed σ_k , then the cost function of the processing delay for each host i to get the computation result from this host can be obtained following M/M/1 queuing model as follows

$$D_{ik}^h(x_{ik}) = \frac{J_k}{\sigma_k - J_k} \quad (4)$$

For network delay, it consists of the time for delivering the raw data from the host to the server plus the time for retrieving the results from the server. Without losing generality, we can assume the volume of the raw data is very larger than that of the results, so the delay of getting the results can be ignored and we only consider the former one. Assume the path P_{ik} between a node i and node k consists of a set of links of l_0, l_1, \dots . Then the cost function between i and k can be represented by a superposition of delays of all links along the path by modeling each link as M/M/1 queue as follows

$$D_{ik}^n(x_{ik}) = \sum_{l \in P_{ik}} \left[\frac{Y_l}{C_l - Y_l} + d_l Y_l \right] \quad (5)$$

where Y_l is the aggregated traffic over link l including x_{ik} , C_l and d_l are the link capacity and propagation delay of link l respectively.

Summing up, the overall cost function for response time between host i and k is

$$D_{ik} = D_{ik}^h + D_{ik}^n \quad (6)$$

Consider the overall cost function of the form

$$\sum_{i \in N} \sum_{k \in S_i} D_{ik}(x_{ik}) \quad (7)$$

The problem is to find a set of job assignments $\{x_{ik}\}$ that minimize the cost function (6) subject to the constraints of equations (1)-(3). More formally, the load balancing problem can be stated as follows

$$\begin{aligned} & \text{minimize } D(x) \\ & \text{subject to } \sum_{k \in S_i} x_{ik} = r_i, \quad \forall i \in N \quad (8) \\ & \quad \quad \quad x_{ik} \geq 0, \quad \forall i \in N, k \in S_i \end{aligned}$$

where

$$\begin{aligned} D(x) &= \sum_{i \in N} \sum_{k \in S_i} D_{ik}(x_{ik}) \\ &= \sum_{i \in N} \sum_{k \in S_i} (D_{ik}^h(x_{ik}) + D_{ik}^n(x_{ik})) \end{aligned} \quad (9)$$

2.2 Necessary Conditions for Optimality

We first calculate the partial derivative of the objective cost function (9) with respect to x_{ik} , which is

$$\frac{\partial D}{\partial x_{ik}}(x) = D_{ik}^h + D_{ik}^n \quad (10)$$

where D_{ik}^h and D_{ik}^n are the derivatives of $D_{ik}^h(x)$ and $D_{ik}^n(x)$ respectively. Equation (10) can be interpreted as the overall response time between host i and server k if the server's processing delay is taken to be the first derivative D_{ik}^h and the network delay to be the first derivative D_{ik}^n .

According to the Kuhn-Tucker theorem(see also [9]), at optimality, a host only assigns its jobs over servers that have the minimum response time given by (10). More specifically, let $x^* = \{x_{ik}^*\}$ be an optimal assignment vector. Then if $x_{ik}^* > 0$ for some $k \in S_i$, we have[9]

$$\frac{\partial D}{\partial x_{ik}}(x^*) \geq \frac{\partial D}{\partial x_{ik}}(x^*), \quad \forall k \in S_i \quad (11)$$

That is, the optimal assignment is *positive* only on servers with the minimum first derivative response time.

3. LOAD BALANCING ALGORITHM AND DISTRIBUTED IMPLEMENTATION

3.1 Gradient Projection Methods

The standard technique to solve the equality constrained optimization problem (8) is the *gradient projection algorithm* [9]. The basic idea is to iteratively adjust the amount of assignment in opposite to the gradient of the objective function and projected to feasible space specified by the constraint functions. In general, the gradient projection algorithm takes the iterative form as follows

$$x^{t+1} = \left[x^t - \alpha \nabla D(x^t) \right]^+, \quad t = 0, 1, \dots \quad (12)$$

where $\alpha > 0$ is the step size at time t , $\nabla D(x^t)$ vector with the (i,k) th element to be the first derivative of $D(x)$ with respect to x_{ik} , which is $\frac{\partial D}{\partial x_{ik}}(x)$ as shown in (10) at time t , and for any vector z , the function $[z]^+$ denotes the projection of z onto the feasible space.

The iterative algorithm in (11) can be expressed individually for each host-server pair as the following

$$x_i^{t+1} = \left[x_i^t - \alpha \nabla D_i(x^t) \right]^+ \quad (13)$$

where $x_i^t = \{x_{ik}^t, k \in S_i\}$ is the assignment vector of host i over all its servers at time t .

3.2 Practical Implementation

The projection algorithm in (13) can be carried out individually by each host according to the locally received cost function information, so it is well suited for distributed implementation. The most straightforward approach is for each host to broadcast the value of the current total aggregated loading J_k to all hosts that select it as the server. Each host then computes the host-dependent cost D_{ik}^h and its derivative. The difficult here, however, is that the network-dependent cost D_{ik}^n can only be implicitly obtained since in general it is not practical to ask intermediate network nodes for detailed traffic information.

Notice that the parameters involved in the projection algorithm are only the first derivative D_{ik}^h and D_{ik}^n , which, as we have discussed, can be interpreted as the response time during the host and the network respectively. For host-dependent response time, it is actually the time for waiting for the computation results, while for network-dependent response time, it is actually the end-to-end delay between the host and server. So we can actually directly measure these two

values and use them to approximate the first derivative of two cost functions.

Let d_k be the measured processing delay at host k and d_{ik} be the network delay between host i and k .

Here d_{ik} can be measured by host i itself, while d_k can be measured and fed back by host k . Host i receiving all these information from all its servers can determine the optimal assignment of its task among these servers using the iterative algorithm as follows

-
-
- (1) $d_{i,\min} \leftarrow \min\{d_{ik} + d_k \mid k \in S_i\}$
 - (2) $k_0 = \arg \min\{d_{ik} + d_k \mid k \in S_i\}$
 - (3) **foreach** $k \in S_i$, **do**
 - (4) $\Delta_{ik} \leftarrow d_{ik} + d_k - d_{i,\min}$;
 - (5) **if** $k \neq k_0$, **do**
 - (6) $x_{ik} \leftarrow \max\{0, x_{ik} - \alpha \times \Delta_{ik}\}$;
 - (7) **end**
 - (8) **end**
 - (9) $x_{ik_0} \leftarrow r_i - \sum_{k \in S_i, k \neq k_0} x_{ik}$
-
-

In this algorithm, k_0 is the index of the server with the minimum total response time $d_{i,\min}$. In step (4), the algorithm computes the difference of the response time between any server and this minimum one. In step (6), it decreases the assignment to each non-minimum server in proportion to this difference and projection to the feasible space. The max projection function guarantees that nonminimum with zero assignment originally always stays at zero. In step (9), it simply assigns the rest task to the server with the minimum response time. It is trivial to prove that the assignment to the minimum response time server is always increased, while the nonminimum servers are always reduced until to zero, eventually only these servers with minimum (and identical) response time have positive assignment as (11) indicates. This guarantees that the assignments approach to optimality eventually.

3.2 Discussion

Unlike traditional distributed computing system, the participating hosts in the peer-to-peer network may be distributed across wide area network, which make the network delay a non-negligible factor. At the same time, computers involved in the p2p computing are normally personal computers that have diverse processing speeds. The processing delay is the major concern for the p2p

computing. Our model incorporates both factors into the same framework. Therefore the load balance problem exhibits distinct features from the traditional distributed computing as well as other delay-insensitive p2p systems. Depending on the relation of the network delay and processing delay, we can envision the following two scenarios;

- If the processing delay is the dominate part of the overall response time, i.e., process delay is far more longer than network delay, our model is close to the traditional load balance problem where the objective is to assign the jobs across the servers according to their processing capability.
- If the network delay is the dominate part of the overall response time, our model is close to the optimal routing problem. The objective here becomes to assign the tasks across the servers to avoid overloading the network connections with longer delay.

Clearly, under this framework, in contrast to the conventional wisdom that a host should choose these physically closet nodes as the servers, which may led to task concentration over these nodes and a higher processing delay is expected. Instead, a tradeoff should be achieved to between the processing delay and network delay, which makes the selection of the server a rather challenging problem.

4. PERFORMANCE EVALUATION

In this section, we evaluate the distributed load balancing scheme proposed in this paper using simulations based on a randomly generated network. The objective is to investigate the convergence of the distributed implementation, the effects of the number of servers for each host on the performance of the scheme, and the behavior of the scheme under different composition of the network delay and processing delay. The network used in the simulations has 100 nodes, among of them, 10 nodes are randomly selected as the hosts which generate the raw data at a rate of 10k bits per seconds. The rest nodes serve as the servers for these hosts. The number of servers for all hosts is the same and the map of host and server is randomly determined.

Fig. 1 illustrates the convergence of the mean response time of these 10 hosts within a 200-seconds run of simulation. In this simulation, each host randomly selects 5 servers and distributes its task evenly across its servers at the beginning, which are far from optimal as shown in the figure: all of them incur higher response time which is the mean value of the response time to all the servers of each host. The response time of all hosts drops quickly within the first 20 seconds of the simulation and converges to the stable value after the simulation is run for about 30 seconds. The result suggests that proposed algorithm can reduce the tasks from the overloaded servers or servers with congested network connections and distribute them to the under-utilized servers or these that have good network conditions.

Fig. 2 compares the mean response time of all hosts under different number of servers for each host. We consider five cases where the number of servers of each hosts varies from 2 to 10. As this figure shows, the hosts incur a very higher response time when they have only two servers for each. This is reasonable because in this case a host has little freedom to adjust the assignment of the task if both of its two servers are overloaded or under severe network conditions. The response time is significantly reduced as the number of the servers of each host is increased to four as well as from four to six. However, the improvement of the response time is slowed down as the number of servers is further increased, and the convergence of the response time has a subtle increment when the number of servers is increases. This simulation suggests that the server number should be chosen at a reasonable value so to achieve the tradeoff between the response time and the convergence time.

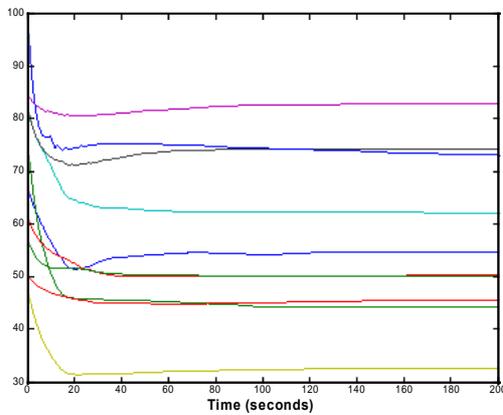


Fig. 1 Convergence of response time vs simulation time

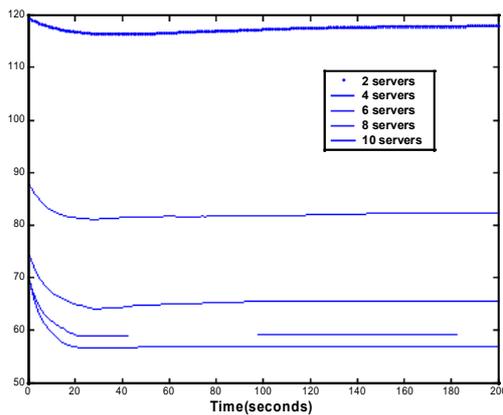


Fig. 2 Response time under different number of servers

As discussed in the previous section, our model tries to capture two factors that affect the performance of p2p computing environments. The proposed scheme may exhibit different properties depending on the composition of the overall response time. In this simulation, we will verify this characteristic using two set of settings. We use the same network topology as above simulations so that the average network delay is

roughly fixed. We then adjust the processing speeds of the nodes so that in the first case, the ratio of the processing delay to the network delay is about 5:1, while in the second setting, this ratio is about 1:5. In Fig. 3 and Fig. 4, we illustrate the throughput (the message processed in bits per seconds) of the host under these two settings respectively. Each figure shows the max, min and mean throughput of all nodes. As shown in Fig.3, since in this case the processing delay is very larger comparing to network delay, the algorithm behaviors like the normal load balancing scheme that can distribute tasks as evenly as possible overall servers. So the max, min and mean value are very closer. While as shown in Fig. 4, here the network delay is comparable to processing delay, so the algorithm considers not only the server load but also the network delay, so the server throughput is not distributed evenly, however, the overall response time still approaches to the optimality not shown here.

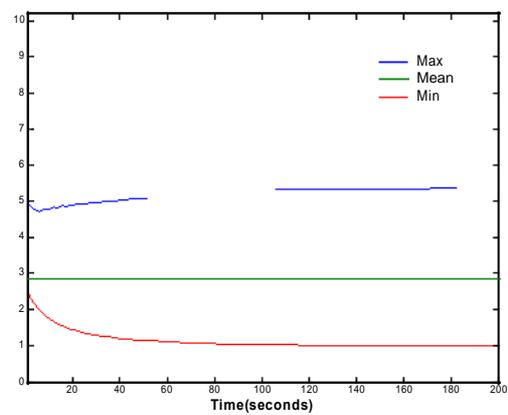


Fig. 3 Throughput under first setting

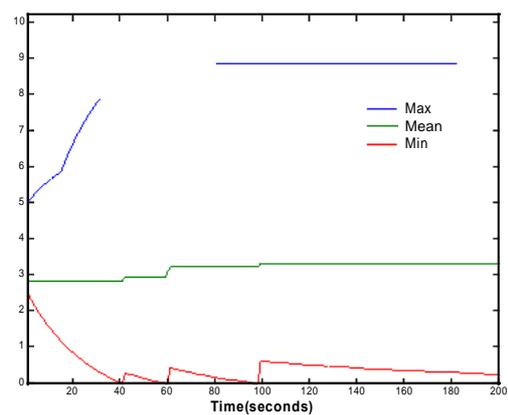


Fig. 4 Throughput under second setting

5. CONCLUSION AND FUTURE WORK

In this paper, we have presented the modeling and solution for the loading balancing problem for a p2p computing system. In this system, the primary challenge is to allocate the tasks overall all participating nodes so that the overall response time is minimized. We model

this as a constrained optimization problem. The model captures the distinct features of p2p computing, that is, the cost involved in the computation includes not only the processing delay, but also the network delay. Leveraging some well studied analysis techniques, we can solve the constrained optimization problem using the gradient projection algorithm, which is well suitable for distributed implementation. Using extensive simulation experiments, we evaluate the proposed scheme in terms of convergence, response time as well as the tradeoff the processing delay and network delay. Experimental results demonstrate our scheme has satisfied performance.

In this study, we have only considered the static situation where the nodes are stable in the network, and a host never changes its servers once selected. It is interesting to investigate the case where the node can join or leave the p2p networks dynamically. A leaving node may broke the ongoing computing, this lead to another problem that a host may have to change its servers for better performance or for failure of the existing servers. It is interesting to investigate the stability of the proposed scheme in this dynamic environment, which will be the focus of the further work.

REFERENCES

- [1] I. Stoica, R. Morris, David Karger, M. Frans Kaashoek, and H. Balakrishnan, "Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications," in *Proc. ACM SIGCOMM*, San Diego, 2001, pp. 149–160.
- [2] A. Rao, K. Lakshminarayanan, S. Surana, R. Karp, and I. Stoica, "Load Balancing in Structured P2P Systems," in *Proc. IPTPS'03*, Feb. 2003.
- [3] J. Byers, J. Considine, and Michael Mitzenmacher, "Simple Load Balancing for Distributed Hash Tables," in *Proc. IPTPS'03*, Feb. 2003.
- [4] B. Godfrey, K. Lakshminarayanan, S. Surana, R. Karp and I. Stoica, "Load Balancing in Dynamic Structured P2P Systems", *Infocom'04*, March, 2004.
- [5] Yingwu Zhu and Yiming Hu. "Towards Efficient Load Balancing in Structured P2P Systems". *IPDPS'04* . April 2004.
- [6] David R. Karger and Matthias Ruhl, "Simple Efficient Load Balancing Algorithms for Peer-to-Peer Systems", *IPTPS'04* . Feb. 2004.
- [7] Subhash Suri, Csaba Toth and Yunhong Zhou. "Uncoordinated Load Balancing and Congestion Games in P2P Systems", *IPTPS'04* . Feb. 2004.
- [8] Dimitri P. Bertsekas and John N. Tsitsiklis, "Parallel and Distributed Computation: Numerical Methods", Athena Scientific, 1997.
- [9] Dimitri P. Bertsekas and R. Gallager, "Data Networks", Printice-Hall Inc. 2nd edition, 1992.

Survey of Security in Grid Services

Jianfang Xiao, Dongdai Lin

The State Key Laboratory of Information Security, Institute of Software, Chinese Academy of Sciences,
Graduate School of the Chinese Academy of Sciences (GSCAS), Beijing 100080, China
Jianfang02@ios.cn

ABSTRACT

This article provides a survey of Security in Grid Services coming from a study of many papers most of which were done by the Grid Forum OGSA-SEC (Open Grid Service Architecture Security) working group, GSI (Grid Security Infrastructure) working group, and Globus Alliance team and other people who contributed to Grid. It describes the best practice in terms of Grid Security Challenges, Grid Security Requirements, and the GT3 (Globus Toolkit version 3) Security Model for OGSA. Most of these were further refined in separate documents.

Keywords: Grid, Security, GSI, Globus

1. INTRODUCTION

The term "Grid" denotes a proposed distributed computing infrastructure for advanced science and engineering [1, 2]. The Grid Security Infrastructure (GSI) was subsequently developed, based on existing standards, to address the unique security requirements that arise in Grid environments, as described in [3, 4, 5].

Research and development efforts within the Grid community have produced protocols, services, and tools that address the challenges arising when we seek to build scalable *virtual organizations* (VOs). Here a virtual organization is defined as a set of individuals or institutions sharing resources and services under a set of rules and policies governing the extent and conditions for that sharing [6].

Controlling access to services through robust security protocols and security policy is paramount to controlling access to VO resources and assets. Thus, authentication mechanisms are required so that the identity of individuals and services can be established, and service providers must implement authorization mechanisms to enforce policy over how each service can be used. The requirement for composition complicates issues of policy enforcement, as one must be able to apply and enforce policy at all levels of composition and to translate policies between levels of composition [6].

Grid computing research has produced security technologies based not on direct interorganizational trust relationship but rather on the use of the VO as a bridge among the entities participating in a particular community or function. The results of this research have been incorporated into a widely used software system called the Globus Toolkit (GT)[3] that uses public key technologies to address issue of single sign-on, delegation [7], and identity mapping, while supporting standardized APIs such as GSS-API [8].

The recent definition of the Open Grid Services Infrastructure specification and other elements of the Open Grid Services Architecture (OGSA) [9] within the Global Grid Forum introduce new challenges and opportunities for Grid security.

Integration of GSI with OGSA enables the use of Web services techniques to express and publish policy [10], allowing applications to determine automatically what security policies and mechanisms are required of them. Implementing security in the form of OGSA services allows those services to be used as needed by applications to meet these requirements [11].

The remainder of this document is structured as follows. First, the following sections present a review of security challenges encountered in grid environments. Then discuss the security requirements. At last the document survey an OGSA security model and the GT security model.

2. SECURITY CHALLENGES IN A GRID ENVIRONMENT [6]

The security challenges faced in a Grid environment can be grouped into three categories. A solution within a given category will often depend on a solution in another category.

2.1 The Integration Challenge

It is unreasonable to expect that a single security technology can be defined that will both address all Grid security challenges and be adopted in every hosting environment. Thus, to be successful, a Grid security architecture needs to step up to the challenge of integrating with existing security architectures and models across platforms and hosting environments. This means that the architecture must be *implementation agnostic*, so that it can be instantiated in terms of any existing security mechanisms (e.g., Kerberos, PKI); *extensible*, so that it can incorporate new security

services as they become available; and *integratable* with existing security services.

2.2 The Interoperability Challenge

Services that traverse multiple domains and hosting environments need to be able to interact with each other, thus introducing the need for interoperability at multiple levels: At the *protocol level*, we require mechanisms that allow domains to exchange messages. This can be achieved via SOAP/HTTP, for example; At the *policy level*, secure interoperability requires that each party be able to specify any policy it may wish in order to engage in a secure conversation—and that policies expressed by different parties can be made mutually comprehensible; At the *identity level*, we require mechanisms for identifying a user from one domain in another domain. This requirement goes beyond the need to define trust relationships and achieve federation between security mechanisms (e.g., from Kerberos tickets to X.509 certificates).

2.3 The Trust Relationship Challenge

Grid service requests can span multiple security domains. Trust relationships among these domains play an important role in the outcome of such end-to-end traversals. A service needs to make its access requirements available to interested client entities, so that they understand how to securely request access to it. Trust establishment may be a one-time activity per session or it may be evaluated dynamically on every request. The dynamic nature of the Grid in some cases can make it impossible to establish trust relationships among sites prior to application execution [4]. Given that the participating domains may have different security infrastructures (e.g., Kerberos, PKI) it is necessary to realize the required trust relationships through some form of federation among the security mechanisms.

The trust relationship problem is made more difficult in a Grid environment by the need to support the dynamic, user-controlled deployment and management of *transient services* [1]. End users create such transient services to perform request-specific tasks, which may involve the execution of user code.

More details about Security Challenges in a Grid Environment occurred in Ref. [6].

3. GRID SECURITY REQUIREMENTS [5, 6]

That the goal and purpose of Grid technologies is to support the sharing and coordinated use of diverse resources in dynamic, distributed VOs: in other words, to enable the creation, from distributed components, of virtual computing systems that are sufficiently integrated to deliver desired qualities of service. Security is one of the characteristics of an

OGSA-compliant component. The basic requirements of an OGSA security model are that security mechanisms be *pluggable* and *discoverable* by a service requestor from a service description. This functionality then allows a service provider to choose from multiple distributed security architectures supported by multiple different vendors and to plug its preferred one(s) into the infrastructure supporting its Grid services.

OGSA security must be seamless from edge of network to application and data servers, and allow the federation of security mechanisms not only at intermediaries, but also on the platforms that host the services being accessed.

A Grid security solution should be based on existing standards wherever possible. Security is an extremely complex problem, with specific solutions incrementally developed over many years by many extremely talented people. Further, the community generally only trusts a particular security solution if it has stood the tests of time and repeated scrutiny.

However, Grid environments have a broad range of security requirements [3, 4]. Unfortunately, no single, existing, standard security solution addresses all of these requirements, though ideally a Grid security solution would extend existing standards.

Grid authentication requirements include: Single sign on, Delegation, Integration with various local security solutions, User-based trust relationships.

Grid requirements for communication protection include: Flexible message protection, Supports various reliable communication protocols, Supports independent data units (IDU).

Grid authorization requirements include: Authorization by stakeholders, restricted delegation.

Specification about the Security Requirements can be seen in Ref. [5, 6].

4. GT2 GRID SECURITY MODEL [11]

We first review briefly the security technologies incorporated in the Globus Toolkit version 2 (GT2) [12] before we survey the OGSA and GT3 security. GT2 includes some services that use a common Grid Security Infrastructure (GSI) [3,4] to provide security functionality.

Diverse site security mechanisms. GSI defines a common credential format based on X.509 identity certificates [13,14] and a common protocol based on transport layer security (TLS [15], SSL [16]).

Dynamic creation of entities and the granting of privileges to those entities. GSI introduces X.509 proxy

certificates [14] that allow a user to assign dynamically a new X.509 identity to an entity and then delegate some subset of their rights to that identity. Users create a proxy certificate by issuing a new X.509 certificate signed using their own credentials instead of involving a CA. This mechanism allows new credentials and identities to be created quickly without the involvement of a traditional administrator.

Dynamic creation and management of overlaid trust domains. The requirement for overlaid trust domains to establish VOs is satisfied by GSI using both proxy certificates and security services such as the Community Authorization Service (CAS) [17]. GSI has an implicit policy that any two entities bearing proxy certificates issued by the same user will inherently trust each other. This policy allows users to create trust domains dynamically by issuing proxy certificates to any services that they wish to collaborate.

In designing GSI we evaluated several related efforts before electing to build on PKI. We noted the following shortcoming in other approaches with respect to Grid security requirements [11]:

Kerberos [18] requires the explicit involvement of site administrators to establish interdomain trust relationships or to create new entities.

The CRISIS wide area security system [19] defines a uniform and scalable security infrastructure for wide area systems but does not address interoperability with local security mechanisms.

Secure Shell (SSH) [20] provides a strong system of authentication and message protection but has no support for translation between different mechanisms or for creation of dynamic entities.

The Legion security model [21] is perhaps the most similar to that of GT2, using X.509 certificates for delegation. However, it lacks mechanisms for creation of dynamic entities.

5. THE GT3 SECURITY MODEL FOR OGSA [11]

We now turn to the question of how Grid security challenges can be addressed within the context of the Open Grid Services Architecture (OGSA) [9], a set of technical specifications that align Grid technologies with emerging Web services technologies [16].

Web services technologies allow software components to be defined in terms of access methods, bindings of these methods to specific communication mechanisms, and mechanisms for discovering relevant services. While particular mechanisms and methods are not prescribed, some mechanisms are emerging as ubiquitous.

OGSA defines standard Web service interfaces and behaviors that add to Web services the concepts of stateful services and secure invocation. These interfaces and behaviors define what is called a "Grid service" and allow users to manage the Grid service's life-cycle, as allowed by policy, and to create sophisticated distributed services. Grid services can define, as part of their interface, service data elements (SDEs) that other entities can (again, subject to policy) query or subscribe to.

OGSA introduces both new opportunities and new challenges for Grid security. Emerging Web services security specifications address the expression of Web service security policy (WS-Policy [10], XACML [22]), standard formats for security token exchange (WS-Security [23], SAML [24]), and standard methods for authentication and establishment of security contexts and trust relationships (WS-SecureConversation [25], WS-trust [26]). These specifications may be exploited to create standard, interoperable methods for these features in Grid security. But they may, in some case, also need to be extended to address the Grid security requirements listed above.

GT3 and its accompanying Grid Security Infrastructure (GSI3) provide the first implementation of OGSA mechanisms. GT3's security model seeks to allow applications and users to operate on the Grid in as seamless and automated a manner as possible. Security mechanisms should not have to be instantiated in an application but instead should be supplied by the surrounding grid infrastructure, allowing the infrastructure to adapt on behalf of the application to meet the applications requirements. The application should need to deal with only application specific policy. GT3 uses the following powerful features of OGSA and Web services security to work toward this goal [11]:

- ①. Cast security functionality as OGSA services to allow them to be located and used as needed by applications.
- ②. Used sophisticated hosting environments to handle security for applications and allow security to adapt without having to change the application.
- ③. Publishes service security policy so that clients can discover dynamically what credentials and mechanisms are needed to establish trust with the service.
- ④. Specifies standards for the exchange of security tokens to allow for interoperability.

5.1 Security as Services

Secure operation in a Grid environment requires that applications and services be capable of supporting a variety of security functionality, such as authentication, authorization, credential conversion, auditing, and delegation. Grid applications need to interact with other applications and services that have a range of security mechanisms and requirements. These mechanisms and

requirements are likely to evolve over time as new mechanisms are developed or policies change. Grid applications must avoid embedding security mechanisms statically in order to adapt to changing requirements.

The OGSA security model casts security functions as OGSA services. This strategy allows well-defined protocols and interfaces to be defined for these services and permits an application to outsource security functionality by using a security services with a particular implementation to fit its current need.

5.2 Hosting Environment

It is not a trivial task to fine and use security services such as those described in the preceding section: in fact, it can require considerable sophistication on the part of the application. Ideally, application developers should not be burdened with the details of this process.

Grid services, like the Web services they leverage, may be built on sophisticated container-based hosting environments such as J2EE or .NET. These hosting environments provide a high level of functionality and allow for much security implementation complexity to be pulled from applications. It is envisioned that most security functionality will be placed in hosting environments, simplifying application development and allowing security functionality to be upgraded independently of applications.

5.3 Publishing of Security Policy

In order to establish trusts, two entities need to be able to find a common set of security mechanisms that both understand. The use of hosting environments and security services, as described previously, enables OGSA applications and services to adapt dynamically and use different security mechanisms. However, an application can select the proper security mechanisms and credentials only if it knows what mechanisms and credentials are acceptable to the service with which it wishes to interact.

The WS-Policy [10] specification and its related specifications define how a Web service can publish its security policy along with its interface specification as part of a WASL document. Such a published policy can express requirements for mechanisms, acceptable trust roots, token formats, and other security parameter. An application wishing to interact with the service can examine this published policy and gather the needed credentials and functionality by contacting appropriate OGSA security services.

5.4 Specified format for Security Tokens

The WS-Security [23], WS-SecureConversation [25], and WS-Trust [26] specifications contain conventions

and formats for the communication of various mechanism specific tokens (e.g., Kerberos tickets and X.509 certificates) inside SOAP envelopes. This enveloping standardizes the protocol for security mechanisms and allows mechanisms to be independent of any application protocol. Hosting environments can recognize security-related messages and route them to an appropriate service for handling, and entities in the Network can recognize whether and how an interaction is secured.

5.5 GT3 Security Advantages

The Grid Security Infrastructure version 3 (GSI 3) of the Globus Toolkit version 3 is an initial implementation of key components of the OGSA security model described above. This implementation has two key advantages over its GT2 predecessor described in Section 4:

- ① Use of WS-Security protocols and standards. GT3 uses SOAP and the Web services security specifications for all of its communications. This allows it to leverage and use standard current and future Web service tools and software.
- ② Tight least-privilege model. In contrast to GT2, the GT3 resource management implementation uses no privileged services. All privileged code is contained in two small, tightly constrained setuid programs.

How these two advantages were implemented in GT3 and more details about Section 4, 5 were described in Ref. [11].

6. CONCLUSIONS

Besides the human factor (Users may become a threat to the security of the systems sometimes), Grid computing present so many security challenges that building a secure Grid environment is a complex task. The use of security technologies (e.g. policy languages, authorization services, encryption mechanisms) alone does not make for a secure system. GT3 implements the emerging OGSA; its GSI implementation (GSI 3) takes advantage of this evolution to improve on the security model used in earlier versions of the toolkit [11]. Its development provides a basis for a variety of future work.

ACKNOWLEDGMENT

This paper was supported by the 863 Project (2003AA144030) and NSFC (90204016).

REFERENCES

- [1] Foster, I., C. Kesselman, and S. Tuecke, "*The Anatomy of the Grid: Enabling Scalable Virtual Organizations*," *International Journal of Supercomputer Applications*, 2001.

- [2] Stevens, R. and e. al., "From the I-WAY to the National Technology Grid," *Communications of the ACM*, vol. 40, pp. 50-61, 1997.
- [3] Butler, R., D. Engert, I. Foster, C. Kesselman, and S. Tuecke, "A National-Scale Authentication Infrastructure," *IEEE Computer*, vol. 33, pp. 60-66, 2000.
- [4] Foster, I., C. Kesselman, G. Tsudik, and S. Tuecke, "A Security Architecture for Computational Grids," presented at Proceedings of the 5th ACM Conference on Computer and Communications Security, 1998.
- [5] S. Tuecke, *Grid Security Infrastructure (GSI) Roadmap*,
http://www.gridforum.org/security/ggfl_2001-03/drafts/draft-ggf-gsi-roadmap-02.doc
- [6] Nataraj Nagaratnam, Philippe Janson, John Dayka, Anthony Nadalin, Frank Siebenlist, Von Welch, Steven Tuecke, and Ian Foster, *Security Architecture for Open Grid Services*, <http://www.globus.org/ogsa/Security/draft-ggf-ogs-a-sec-arch-01.pdf>.
- [7] Gasser, M. and McDermott, E., *An Architecture for Practical Delegation in a Distributed System*. Proc. 1990 IEEE Symposium on Research in Security and Privacy, 1990, IEEE press, 20-30.
- [8] Linn, J. *Generic Security service Application Program Interface, Version 2*. INTERNET RFC 2078, 1997.
- [9] Foster, I. And Kesselman, C. Nick, J. and Tuecke, S. *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration*, Globus Project, 2002. <http://www.globus.org/research/papers/ogsa.pdf>.
- [10] BEA, IBM, Microsoft and SAP. *Web Services Policy Language (WS-Policy)*, 2002.
- [11] Von Welch, Frank Siebenlist, Ian Foster, John Bresnahan, Karl Czajkowski, Jarek Gawor, Carl Kesselman, Sam Meder, Laura Pearlman, and Steven Tuecke, *Security for Grid Services*, <http://www.globus.org/Security/GSI3/GT3-Security-HPDC.pdf>
- [12] Foster, I. And Kesselman, C. *Globus: A Toolkit-Based Grid Architecture*. Foster, I. And Kesselman, C. eds. *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann, 1999, 259-278.
- [13] CCITT Recommendation X.509: *The Directory – Authentication framework*. 1998.
- [14] Tuecke, S., Engert, D. Foster, I., Thompson, M., Peatman, L. and Kesselman, C. *Internet X.509 Public Key Infrastructure Proxy Certificate Profile*, IETF, 2001;
- [15] Dierks, T. and Allen, C. *The TLS Protocol Version 1.0*, IETF, 1999. <http://www.ietf.org/rfc/rfc2246.txt>.
- [16] Graham, S., Simeonov, S., Boubez, T., Daniels, G., Davis, D., Nakamura, Y. and Neyama, R. *Building Web Services with Java: Making Sense of XML, SOAP, WSDL, and UDDI*. Sams, 2001.
- [17] Pearlman, L., Welch, V., Foster, I., Kesselman, C. and Tuecke, S., *A Community Authorization Service for Group Collaboration*. IEEE 3rd International Workshop on policies for Distributed Systems and Networks, 2002;
- [18] Neuman, B. C. and Ts'o, T. *Kerberos: An Authentication Service for Computer Networks*. *IEEE Communications Magazine*, 32(9). 33-88. 1994.
- [19] Belani, E., Vahdat, A., Anderson, T. and Dahlin, M. *The CRISE Wide Area Security Architecture*. 8th Usenix UNIX Security Symposium, 1998.
- [20] *OpenSSH*, <http://www.openssh.com>, 2003.
- [21] Humphrey, M., Knabe, F., Ferrari, A. and Grimshaw, A., *Accountability and Control of Process creation in Metasystems*. 2000 *Network and Distributed System Security Symposium*, 2000.
- [22] *eXtensible Access Control Markup Language (XACML) 1.0 Specification*, OASIS, February 2003. <http://www.oasis-open.org/committees/xacml/>
- [23] IBM, Microsoft, *RSA Security and VeriSign. Web Services Security Language (WS-Security)*, 2002.
- [24] *Security Assertion Markup Language (SAML) 1.0 Specification*, OASIS, November 2002. <http://www.oasis-open.org/committees/security/>
- [25] IBM, Microsoft, *RSA Security and VeriSign. Web Services Secure Conversation Language (WS-SecureConversation) Version 1.0*, 2002.
- [26] IBM, Microsoft, *RSA Security and VeriSign. Web Services Trust Language (WS-Trust)*, 2002.

A Study of Internet Book Sale

Gaojia Jin, Takasi Nemoto

Department of Business Administration, The Graduate School of Meiji University, Tokyo 101-8301, Japan
{Gaojia Jin}cg44002@kisc.meiji.ac.jp

ABSTRACT

This report has analyzed the business model of KINOKUNIA online book service, by comparing the service of KINOKUNIA online book shop with the other four big online book shop in Japan, some advantages and disadvantages of KINOKUNIA have been found. By analyzing and comparing the service between the services of different online book shop, a more competitive advantage e-business model with quicker response and lower cost was proposed.

Keywords: book sale, e-business model, competitive advantage, net-sale

1. INTRODUCTION

While the term e-commerce refers to all online transactions, B2C stands for "business-to-consumer" and applies to any business or organization that sells its products or services to consumers over the Internet for their own use. When most people think of B2C e-commerce, they think of Amazon.com, the online bookseller that launched its site in 1995 and quickly took on the nation's major retailers. However, in addition to online retailers, B2C has grown to include services such as online banking, travel services, online auctions, health information and real estate sites.

In recent years, with the rapid development of communication technology, the Internet shopping has been more and more popular within the consumer behaviors. The internet business has got a significant development in Japan from 1998. This new kind of selling has supplied multi-choices to the customers before they made their decisions of purchase.

2. THE SITUATION OF INTERNET BOOK SALE

2.1 Background

Because of the widely use of internet in these years, there were significant changes happened on business environment, this new low cost, high efficiency communication method made the quick transfer of information of demand and supply be possible, and which formed a new business mode--- e-business. E-business can be described as "doing business by an electronic mode", much of the interest in e-business in the 1990s was driven by the belief that it would be the source of exciting new revenue opportunities both for existing businesses and for new start-up businesses. Both the business-to-consumer (B2C) and business-to-business(B2B) models proliferated as traditional companies and new e-commerce companies entered the online marketplace. The book sale by internet is a typical model of business-to-consumer. The revenue opportunities these models represent are almost endless. Nevertheless it is possible to understand them as falling into several broad categories driven by the nature of online commerce itself:

- ò Enhanced access to customers
- ò Enhanced sales to existing customers
- ò Electronically enabled products

From 1998, E-business has got a significant development in Japan, one of the pertinent causes is the increasing amount of internet users, until Feb. 2004, there were 62,844,000 internet users in Japan, and in the foreseeable future, this amount would increase to 67,000,000 in Dec. 2004, which would be the third large amount of internet users in the world. [6] (Fig 1)

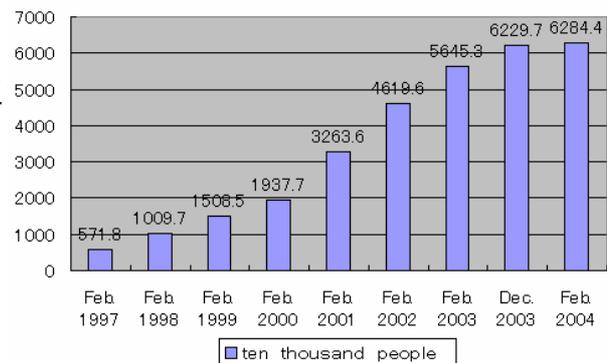


Fig. 1, the number of internet users in Japan from 1997 to 2004 Source: Internet White Paper 2004 pp46 by author

With the widely use of internet, more and more people have accepted the new internet sale mode, in Jun. 2003, vagabond and eye bridge company has investigated 1000 internet users for internet shopping, more than 70% of them have used internet shopping services, and in all the things they bought on internet, most of them were low price things, books are the most regular bought things which was 13.3% in all.[13]

2.2 Today's online book sale in Japan

The online book sale has begun from 1995 in Japan, at that time, one of the biggest book-shop---MARUZEN has built a database of 3.6 million books which were written not only in Japanese language but also in foreign languages including Chinese, English, and French etc,

the services of this online bookshop also included credit payment and home delivery. The MARUZEN online bookshop was the first big online book-shop in Japan. In the following several years, nearly all the biggest book shop in Japan have built their own online website, including KINOKUNIYA, SANSEIDO, BUNKYODO. Because of the high profit of online book sale, some publishers and big book-agents also joined in this kind of service, even some big delivery company, for example, the YAMATO transport service group has built its own book service website. Now, there are about 100 online book shop in Japan, about 10 of them are leading the way. The classification of them is (1) Combination of actual bookshop and online bookshop--- normally, the online bookshop is built based on actual bookshop for purpose of increasing sales and supplying better services, such as KINOKUNIYA, MARUZEN, SANSEIDO, and BUNKYODO; (2) Simplex virtual bookshop--- only online bookshop without actual bookshop, such as ESBOOKS, RAKUDEN, and BKI; (3) Books selling on a commission basis virtual bookshop--- this kind of bookshop is run by several shareholders, including book publishers, sale agents and even delivery companies, such as E-HON, HONYA-TOWN, and YAMATO.[5]

The common ground of all the online book shops is they all have a public database on their website and their customers can search and order their books online and their books would be delivered to their home. The difference between different online book shops is the results got from the database are different and the different payment, supply and delivery methods chosen and time. (Fig. 2)

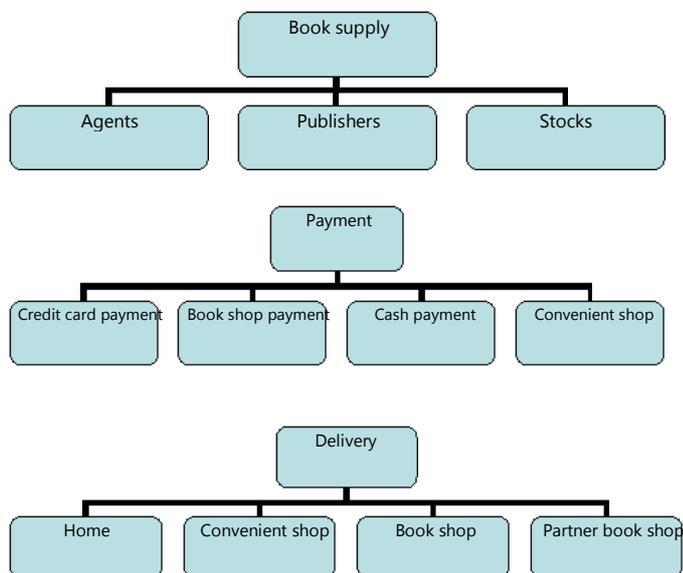


Fig. 2, Classification of Payment, Supply and Delivery

3. THE ONLINE BOOK SHOP OF KINOKUNIA ANALYSIS

3.1 General situation of KINOKUNIA

KINOKUNIYA is one of the biggest online bookshops

in Japan. In 1996, it has launched its virtual bookshop--- KINOKUNIYA BOOKWEB. It was a new online ordering and search service for Japanese books and foreign language books for individual customers. Even in the situation of drastic horizontal competition, it got more than 200 million sales per month in average. Now, as a representative internet shop in Japan, it got a significant development and more and more comprehensive concern.

The online service can only be used by members, although non-members could use the service too, but the search results and display pages are limited by the server, and they couldn't use some special only for member services, so to use all functions of the online book shop of KINOKUNIA, the best way is to be member of it. It is easy to become a member of it, you can do it online, before 2002, it charged 1,500 Yen to become a member, and now, it is free. After you become a member of KINOKUNIA, you can use the huge database of the book shop to choose your favorite books either in Japanese or in foreign languages. The database includes 1.85 million Japanese books and 2.85 million foreign books, which is one the biggest database in the world, under each book, you can find the details of it, including the contents, the author, publisher, and the stock. In addition, KINOKUNIA online book shop has supplied recommend service to members on their own homepages, you can use key-words, author or theme to get your results from the server. For its good quality services and easy use functions, the book-web of KINOKUNIA has got a lot of encouragements in Japan.

3.2 The processes of online book sale by KINOKUNIA

The KINOKUNIA online book shop has limited some of its services to be used by members only, so the first step for customer to use the online service is to become a member, in the registration step, the customer was required to provide name, home address, contact number, credit card details, etc., after that, the customer would get a KINOKUNIA ID and password, by using this ID and password, the customer can use the online book shop services for search and order books.

After logging on to the homepage of KINOKUNIA, the member can use the database of the online book shop to search the books they want, the result would present the details of the books including the author, content, publisher, edition and price, the customer can also choose the nearest branch to order his book. After the online book shop received the order from the customer, if there are stocks in the shop, the online book shop will give a answer to the customer through its Quick Response System and post the books to the customer, if no stock in the shop, the online book shop would order the books from the publisher or agent, and the customer would get a mail to tell them the situation of his order. The KINOKUNIA online book shop would send the

books to the customer by its specially assigned delivery company. During the process of delivery, the customer can get some auto mail from the online book shop to tell them the situation of his order, and the data when books are delivered and when the books can arrive to the customer, normally, it would take 4 days for Japanese books and 7days for foreign books.

The payment methods can be chosen by the customer, you can pay by credit card when you order the books online, and for using this payment, the online book shop would charge 380 Yen for delivery, or you can choose to pay when the books arrive, it would charge 200 Yen by the delivery company. (Fig. 3).

3.3 The advantages of online book shop of KINOKUNIA

A. Its particular virtual bookshelves that make the customers have a feeling of searching in a real bookshop, some book lovers like to choose their books just before the real bookshelves and which give them a special feeling of reading, to satisfy the need of this kind of customers, KINOKUNIA online book shop has set its virtual bookshelves and by this setting, it can give the customers a feeling like choose books in a real book shop.

B. Huge book database that include 1.85 million Japanese books and 2.85 million foreign books; detailed book information introduction; convenient delivery, the customer can choose the nearest branch and get books in the shortest time, all these services can supply more convenience to the customers, and easier for customer to choose their favorite books.

3.4 The disadvantages of online book shop of KINOKUNIA

A. No bonus system, some of the online book services have a kind of system to return profit to customer, for example, the point return system, the regular customers can get bonus points when they purchase the things from the website, and they can use their points as money on this website. Although it is not the necessary condition for a customer to choose the service, but some customers would be affected by this kind of system, if this kind of bonus system was used by KINOKUNIA, the sales would have an increase. In the modern management, the bonus system is an important method to attract new customers and keep the regular customers. To increase the sales, some promotion methods should be used.

B. The extra cost of delivery of the books, the delivery cost 380 Yen per time, the extra cost would affect the customers who only buy one or two books each time, some of them would choose to the real book shop instead of shopping online, to reduce this kind of customer's burden, the online book shop could charge by each book instead of charge by delivery time.

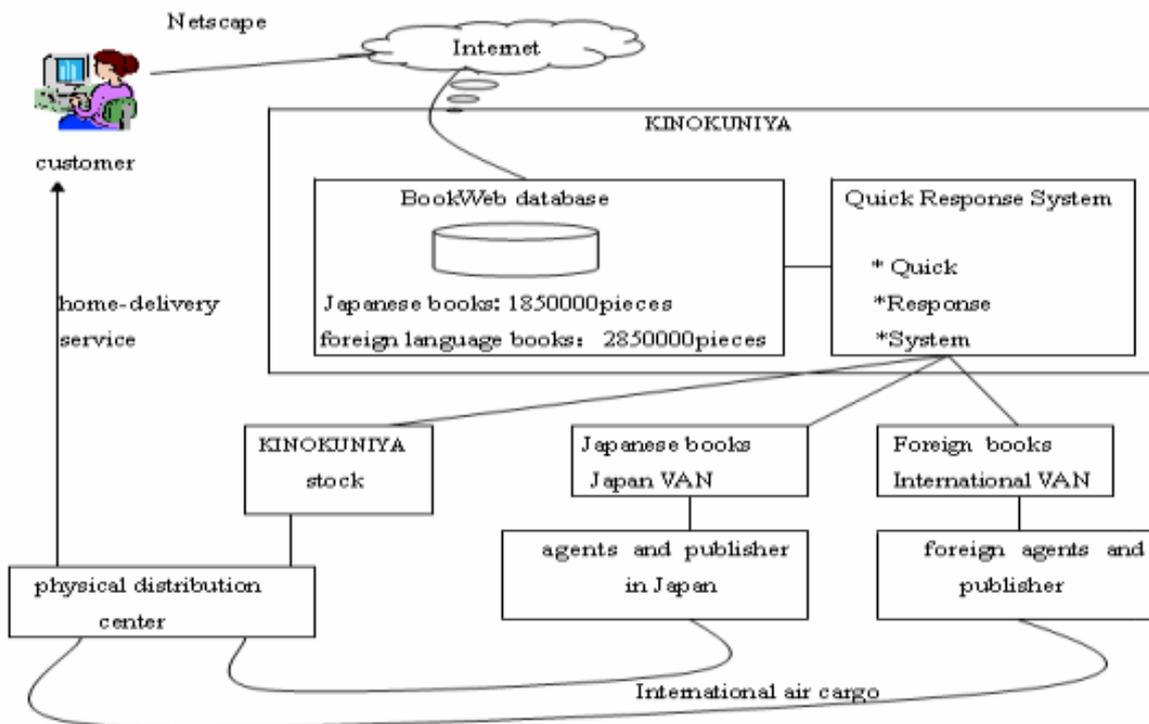


Fig. 3, online book sale process (Source: <http://bookweb.kinokuniya.co.jp> by author)

Table 1, Comparison of services

○ : enable × : disable

store	settlement				delivery place			book review	point
	credit	cash on delivery	over the counter	etc	home-delivery service	foreign	premise		
bookservice	○	○	×	×	○	○	×	×	×
honya-town	○	○	○	×	○	×	○	×	×
esbooks	○	○	○	○	○	×	○	○	×
jbook	○	○	×	○	○	×	×	○	○
bookweb	○	○	○	×	○	○	○	×	×

Source: <http://hon-michi.net/service.htm>

4 THE COMPETITIVE ADVANTAGE E-BUSINESS MODEL

The business model could be affected by the customers shopping requirements; Japan is a country, customers have a lot of opportunities to get their needs by real shops, so consider of this situation of Japan a more competitive e-business model is needed for the internet shops. To satisfy the need of customers, we need to know their concerns about the services, the following figure is an investigation by Myvoice. Com for the concerns when the customers choosing to buy book from online book shop: from this figure, we can see that 33% of customers paid attention to the delivery fee were the categories, book review, delivery time, bonus points, and delivery places, etc.(Fig 4)

book shop of a group of publishers and book agents, the benefit of purchase books there is no charges for the delivery from the nearest branch of its group;

BOOKSERVICE is a online book shop of YAMATO transport company, the advantage of it is the BOOKSERVICE can use YAMATO company's transport to deliver the books and which save time of the delivery places to the customers, they can choose the most convenient place to pick up the books.[12]

By analyzing and comparing the service of KINOKUNIYA BOOKWEB and the other 4 online bookshops in Japan, a competitive advantage e-business model with quicker response, lower cost and more convenient delivery choices was proposed.

The major challenges of B2C online book sale are:

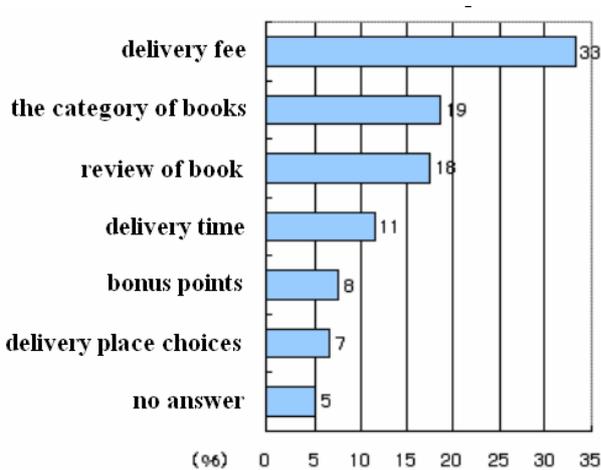


Fig 4, the concerns of customers about services
Source:<http://www.myvoice.co.jp/voice/enquete/2203/index.html>

Getting browsers to buy things — your online book shop site cannot live on traffic alone. Getting visitors to the site is only half the battle. Whether they buy something is what determines if you win. The so-called conversion rate for B2C e-business sites is still fairly low. (Boston-based Yankee Group said in November 2000 that the average rate was 1 percent.) Some ways to boost your conversion rate include improving navigation, simplifying checkout process (such as one-step checkout and easily replaced passwords), and sending out e-mails with special offers. For the online book shop, the best way to attract customers is to make your database user friendly, which means all customers can use your website easily and conveniently.

Building customer loyalty — with so many online book shop sites out there, how can you build a strong relationship with customers? Here are some tips:

By consider of the concerns of the customers in the business management, the biggest online book shops have supply their own special services to their customers, the following table is the services comparison between each online book shop.(Table 1)

Focus on personalization: a wide array of software packages are available to help the online book shops create unique boutiques that target specific customers. For example, American Airlines has personalized its website so that business fliers view it as a business airline and leisure travelers see it as a vacation site. Amazon, which built its own personalization and customer relationship management (CRM) systems, is

Each online book shop has its own characteristics; JBOOK is the online book shop of BUNKYODO book shop, the customer can get 3% bonus points when purchase books online; HONYA-TOWN is the online

well known for its ability to recognize customers' what they concern and make changes to satisfy the needs individual preferences.

2. Create an easy-to-use customer service applicationthe online service. After all, good customer relationships Providing just an e-mail address can be frustrating toare at the heart of business success. customers with questions. Live chat or, at the very least, a phone number will help. For the online book shop, the customers should get quick respond for their questions, so the customer service is important to the online book shop.

5 CONCLUSION

3. Focus on making your site easy to use. Fulfillment preparation to the new chance for their promotion. — E-commerce has increased the focus on customer satisfaction and delivery fulfillment. One cautionary tale The competitive advantage e-business model can help is Toys "R" Us' holiday debacle in 1999, when the online book shop use technology and human fulfillment problems caused some Christmas orders to dresources to gain insight into the behavior of customers delivered late. Since then, companies have spent billions and the value of those customers. If it works as hoped, to improve their logistical systems in order to guaranteethe online business can: on-time delivery. Providing instant gratification for ò provide better customer service customers still isn't easy, but successful B2C ò make book shop branches more efficient e-commerce operations are finding that fulfillmentò cross sell products more effectively headaches can be eased with increased focus andò simplify marketing and sales processes investment in supply chain and logistical technologies. ò discover new customers ò increase customer revenues

In the interest to increase the online book sales, some steps could be taken. First, know your customers, all of them, for the purpose of attracting new customers and keeping the existing customers, the online bookshop should consider more on the sight of customers. Such as both members and non-members can purchase the book on the web site freely, but the members can get some bonus points or discount; the places of book delivery can be chosen by the customers, they may choose from their nearest branches of KINOKUNIYA BOOKWEB, and the cost of different deliveries are different; the BOOKWEB can get the book orders directly from the publisher instead of the agents which can lower the purchase price of the books; more payment methods should be added, such as paying by mail or by BOOKWEB purchase recharge card.

Second, develop IT strategies for specific customer needs, when the customers trying to buy books online they want their orders to be respond as quick as possible, which means they should get the books they ordered quickly, and the payment method they choose should be safe, so that needs the IT support of the online book shop by using the correct IT strategies, the orders of the customers can be deal quickly, safely and correctly.

Third, more communications with customers, the management team should know what the customers need,

REFERENCES

- [1] Asa, Anderson, e Business: Way to Improve All Business Practices Through Network-Enables Technology, Toyo Keizai Inc.,2000
- [2] Junji, Ushiguchi, "E-Commerce and Libraries' Future E-Commerce in Book Trade-experience of Kinokuniya Book Web", The Journal of Information Science and Technology Association,Vol.51,No.1, pp22-26,2001
- [3] Katuyuki, Sugiyama, Internet Marketing Handbook, Tokyo: Douyuukan, 2003
- [4] Tom, Vassos, Strategic Internet Marketing, Tokyo: Asuki, 1998
- [5] Wataru Hoshino,"The Current Situation of Online Bookstores in Japan", Publishing Studies, Vol.32, pp167-172,2001
- [6] Internet White Paper 2004, Internet Association Japan, 2004
- [7] <http://bookweb.kinokuniya.co.jp> ,Jul.16,2004
- [8] <http://www.fri.fujitsu.com/hypertext/fri/cyber>,Sep.16,2004
- [9] <http://www.hon-michi.net/soryo.htm>,Sep.16,2004
- [10] <http://www.hon-michi.net/service.htm>,Sep.16,2004
- [11] <http://www.myvoice.co.jp/voice/enquete/2203/index.html>,Sep.16,2004
- [12] <http://www.hon-michi.net/s.shtml>,Sep.16,2004
- [13] <http://www.research-plus.net>/Sep.12, 2004

Achievements and Future Trends of E-Government Service – Implications from the SARS Outbreak

Zichao Li¹

¹ School of Industrial Engineering, Georgia Institute of Technology, Atlanta, GA 30318
zli@isye.gatech.edu

ABSTRACT

Rapid advances in technology and the advent of the Internet have redefined public expectation of the government and its services. With the great support from Hong Kong and Singapore government, e-government system has achieved temporary success by offering some everyday services through virtual Internet. During last year's Severe Acute Respiratory Syndrome (SARS) outbreak, e-governments in Hong Kong and Singapore ensured the proper running of government services despite of massive physical restrictions caused by SARS outbreak. E-government started to show its value through continuous development in the past decade. The idea of "virtual government" implementation had been proved in some extent. In the meantime, SARS outbreak introduced new challenges to e-government in dealing with abrupt events. The paper will present the new trends and dynamics in e-government development enlightened by feedbacks from fighting with SARS outbreak. The paper will also provide assessment for those functional specifications raised under the SARS outbreak.

Keywords: E-government, CRM, information sharing, e-procurement, SARS

1. TITLE OF THE PAPER

E-government has been broadly defined as the production and delivery of government services through IT applications [6]. Government use of technology, particularly web-based internet applications, to enhance citizen access to government services and to enable citizens to make online transaction is called e-government [9]. E-government has been developed for the delivery of government services using information technology for many years in different countries. E-Government initiative is not just about using IT for government services; it is also an initiative to increase government transparency and increasing governmental efficiency.

E-government includes several types of framework and it behaves differently from country to country. There are three major domains of e-Government – E-Administration (process improvement), E-Services (connecting citizen) and E-Society (building interaction with and within civil society) [3]. Most literature points out the need for e-government initiatives is to reach out to four core groups, in a bid to enhance the total experience level – citizens, business community, government employees and government agencies itself [2][6]. These give rise to the four major category of e-government services, commonly known as Government-to-Citizen (G2C), Government-to-Business (G2B), Government-to-Employees (G2E) and Government-to-Government (G2G).

In this paper, we will first review the various stages towards the e-government maturity, followed by presenting the achievement of the e-government service performance in Hong Kong and Singapore. Other than successful achievements, the paper will review

e-government's role during the SARS outbreak and hence investigate the functional improvement in each of the four major sub-categories. The objective of the analysis and evaluation is the identification of potential improvement area arises from SARS outbreak at growth stage of e-government services globally.

2. PAST ACHIEVEMENTS

Through years of evolvments, developments in G2B, G2C, G2E and G2G services has resulted in two major benefits – efficient governments and innovative governments. E-Government also provides a strategy for addressing the challenges most governments face – such as reduction of operation costs, retention and training of staff and responsiveness to change and for creating an environment for transformation of government activities by the application of E-business methods throughout the public sector E-Government is also positioned by researchers as a promising approach for reducing bureaucracy in Government organizations and increasing citizens' access to the quality services [2].

Past literatures have addressed the achievement of e-government system towards maturity [6]. These stages are summarized as the evolvment in terms of interaction, technical sophistication and service hospitality.

2.1 Information online stage

This stage is the most basic form of e-government marked by the one-way communication between government and users. When Internet gradually goes into everyday life in the mid-90s, governments in developed countries choose to post information or data

on the Internet for public reference. Although this implementation doesn't involve much technical sophistication, it is an effective way for public to be conscious about how government runs and reduces the deviation of information passing between government and users.

Singapore's e-government initiative can be traced back to the early 80s, with the launch of the Civil Service Computerization Program (CSCP). This initiative was conceived as a clear direction of turning government into a world class exploiter of IT and stimulates the IT penetration in Singapore Public Services. Since CSCP's launch in 1981, it had progressively advanced and evolved with the shifting trends in the technological, business and social climate to bring about stimulating changes to the way that government interacting with the public. In the mid 90s, Internet was introduced as a new delivery channel providing information. The Singapore ONE broadband initiative stimulated the delivery of interactive, multimedia applications and services to homes, businesses and schools throughout Singapore.

Hong Kong government started uploading governmental information to their website since the promulgation of the first Digital 21 Information Technology Strategy in 1998. By bring e-government into the agenda of all government bureaus, citizens has gain more accessibility and awareness to the government information and services. Weather forecast and government announcement information are initial service offering in Hong Kong at this stage.

2.2 Simple Communication Stage

This stage involves two-way communication as an interactive mode between Government and constituents [6]. In terms of technical effort, government has used email and feedback panel to answer constituents' request. Most idea of further e-government development is generated from this stage, where government archives constituents' request

In Hong Kong, the average monthly number of page views of all government websites amounts to 155 million, an 80% increase in the Year 2000. Over 1500 government forms can be downloaded online and 400 e-forms are available for completion and submission online. Requests for access to government information are available for submission online.

Singapore's ground breaking e-government service in this stage is the One Stop Change of Address Reporting Services (OSCARS).

2.3 Transactional Stage

The true interaction between government and citizens is enabled by online execution of governmental services

and subsequent financial transactions. The milestone of this stage is connecting the government system with users through web interfaces and allowing citizens with Government electronically anytime – saving hours of paperwork. On the other hand, citizens saved time and inconveniences of traveling to a government agency and queuing time as well.

Typical achievement of this stage in Singapore is the development of the electronic import/export trade documentation system – TradeNet. In the meantime, Hong Kong government opted to put the tax reporting service online as well.

2.4 Service Integration Stage or Seamless Stage

After the development at the transaction stage, most government agencies put up their services on the Internet. However, there may also result in chaos with less integrated service. Some even don't run on the same platform and don't share a common look and feel. Under this context, government attempts to integrate various government services horizontally (integrate with other governmental agencies) and vertically (integrate similar services together) for the enhancement of efficiency, user friendliness, and effectiveness [6]. The development at this stage requires tremendous amount of time and resources to integrate online and back-office systems. Citizens hope to have a uniformed way in accessing all governmental services. Through function re-assessment and combination, the ultimate goal of this stage is to enable user to accomplish all services from a single access point [5].

3. EFFECTIVENESS OF E-GOVERNMENT SYSTEM DURING SARS OUTBREAK

E-government can be explained as the public-sector adoption of e-commerce or e-business. From previous literature [2][6][7]. There are five key principles summarized to measure e-government system's effectiveness:

- *Increased general accessibility.* In the first place, as an e-commerce or an e-business application, it should have. Most literature points out the need for e-government initiatives to reach out to three core groups, in a bid to enhance the total experience level – citizens (or individuals), business community and with itself (interactions within the government).

- *Pure digital processing.* In the meantime, as an IT system, e-government system should have embodied the principle of "put everything online and do everything online", i.e. successful e-government system should have an aggressive approach to move of information from paper form to digital form and completely replace the paper form processing.

- *Dynamic Capabilities.* It should have dynamic capabilities, which is the e-Government's ability to integrate, build and reconfigure internal and external

organization skills, resources, and functional competencies within a changing environment.

3.1 SARS Outbreak

SARS outbreak in 2003 originated from China. It was previously unrecognized to be potentially fatal. The contagious *Coronavrius* infection emerged in East Asia with subsequent global spread. SARS officially hit the global headline on March 12th, 2003 through a global alert by World Health Organization (WHO). Hong Kong and Singapore both were largely affected.

In Hong Kong, in the five-month intensive fighting with SARS from February to June, there are 299 reported deaths. During March 2003, health officials in Hong Kong and Singapore have implemented quarantine and isolation measures to try and limit the spread of SARS. Temperature screening and health declaration by travelers have been a standard practice at all cross-boundary and international immigration checkpoints. Both two governments have founded hard to management the large scale of data sharing between governmental agencies, hospitals and civil service personnel. E-government system could further improve its functional capabilities upon the challenge encountered during the SARS outbreak.

In the following section, the paper will illustrate the potential trend in each subsection of the e-government arise during the SARS outbreak.

3.2 G2E & G2C Service Under SARS outbreak

One of the initial motivations of e-government service is to simplify its citizen services and decreases the number of paper processing which is costly and difficult to process. G2E services can be seen as an internal B2C services which is nothing special in the Internet era. The G2C services offered by Hong Kong and Singapore government are quite comprehensive. Common functions like tax paying, vehicle license application, identity card application and facility booking has all been available in Singapore and Hong Kong. In this section I will focus on the contrast between people's different attitude towards e-Government service before and after the SARS outbreak according to statistics conducted by Hong Kong and Singapore government.

Although SARS has brought disastrous effect o the regional economy and social order, it stimulates the public adoption to e-government system. During SARS outbreaks, people in Hong Kong normally choose to stay at home and wouldn't be able to access to government agencies through physical means. They gradually exploring other approaches to accomplish those tasks which used to available through physical contact. As a result, the percentage of public utility bills paid through electronic means increase from 29% a year ago to 53% at the end of year 2003. Before the SARS

outbreak in December 2002, 39% of the income tax declaration is done through the e-government system in Singapore to pay income tax, however, this figure has increased to 67% after the SARS outbreak in 2003. The e-government system in Hong Kong and Singapore proves its effectiveness in increasing accessibility during the SARS outbreak.

When citizens are getting used to e-government services, ensuring users' loyalty to e-government services becomes increasing important. Moreover, in order to make e-government system to be more effective, it should leverage on private sector's business ideas and innovation so as to improve services to customers [4]. Therefore, Customer Relationship Management (CRM) features have become increasingly popular among e-government services.

During the SARS outbreak, citizens realized that most e-government services aren't specific to user itself, e.g. they cannot find the SARS regulation pertinent to their own interest. The current e-government B2C services are more uniform, where different users cannot enjoy customized services. However, the core idea of CRM theory is to provide more customer-centric and tailor-made services to citizens and business [8].

Common e-business application like online bookstore usually keeps a profile for returning users and offer personalized search according to the user's profile. Hong Kong and Singapore have the world's highest mobile phone penetration rates, with 3G service impending in the next year. Leveraging the mobile phone concentration would stimulate the e-government to offer more personalized system. The most tempting feature of the m-commerce system is offering the location-based services. For example, users could receiving the booking of nearby public sports facilities through there mobile phone. This further motivates the e-government system to garner the concept of CRM and offer the service to the user who really needs the service.

To implement such a customer orientated e-government system, we can leverage on the current CRM framework. The critical success factor for most CRM system includes:

Knowledge of Customer: E-government system needs to gain users knowledge such as preference, and then use business intelligence tools such as data mining, processing and reporting to form a generic profile. Further possible information to be collected includes customer's demographic information, service history and all previous transaction with the e-government system [1]. When dealing with mobile users, while methods like forms and cookies can be used to collect data, there must be new ways to consolidate customer data through their mobile device with limited storage space. Moreover, the mobile ways of collecting data

must be able to conform to the database formats using web interface.

Capturing and increasing added value. E-government system must continuously innovate to stay ahead of users to make the service more superior after each user transaction. For example, if a user access a sports facility where he used to use his mobile phone to book, e-government system could provide better booking information.

Customer Interaction: E-government system could provide customized user interface and periodically seeking online feedback from users. E-government system could also integrate the real-time online service answering features.

Customer Acceptance: Before a customer-orientated e-government initiative to put into practice, developers need to decide the best mean to implement those functional requirements. For a particular service, they need to decide whether needs to use a graphical rich interface or pure text interface. For integration of mobile service, they need to decide the technology (WAP, SMS, GPRS etc) by considering which technology is more appropriate to reach all users.

The key value proposition for G2C initiatives rests on providing electronic services to citizens that are integrated and user-centric, on a 24*7 timeframe, anytime, anywhere, and through a range of channels.

3.3 G2B Services

G2C and G2E services are e-services for individuals with limited transactions and technical sophistications. However, G2B & G2G are the extensions that generate most potential. A centralized G2B e-procurement system, in the framework of e-government, reduces the cost of issuing invoices and other administrative work for low-value, high-volume goods. It also simplifies the CRM in the government sector. However, the largest potential saving originates from lower prices due to centralized spending. This is motivated by the lack of organization in governmental purchase of SARS-related items like thermometer and masks during the SARS period. The critical success factors (CSFs) of such e-procurement system includes:

- Efficient processes without excessive idle times
- Existence of monitoring and evaluation systems that permit the continuous improvement of the process
- Adequate training of the employees in order to enable them taking advantage of the new system.

The above CSFs should be coupled by a set of KPIs that should be monitored as used in enterprise level e-procurement system:

- Tender lead times
- Percentage of governmental purchasing with accepted quality (in accordance with the predefined

technical specifications).

- Productivity of resources
- Budget accuracy (deviation of the actual governmental purchasing compared with the related budget).
- Cost (including all the relevant cost categories, together with the cost of the purchased goods or services).

Another possible trend in e-government is public-private partnerships. Firstly, having business involvement brings about competitive pressures in the e-government journey, in line with Osborne and Gaebler's principles of reinventing government [7]. Secondly involvement of business brings in the required expertise from the market to turn "e" quickly as it is not possible for government agencies to train staff overnight or attract large amount of such talents from the market easily.

3.4 G2G Services

G2G service is very limited in both e-governments of Singapore and Hong Kong, where different government agencies could integrate their services or form partnerships to offer better service and information transparency between agencies as well. By performing vertical and horizontal integration, an e-government system could leverage the strength of e-services provided by different government agencies.

During the SARS outbreak, the *Contact Tracing Information System* implemented in Singapore has been a good start of G2G system where different governmental agencies provide and demand different domains of information. The Ministry of Health, Immigration Checking Authority, Hospitals, Ministry of Education are sharing the same information knowledge base but perform different tasks during the combating with SARS.

Inter-governmental information sharing could be an inflexible process nowadays. During the SARS outbreak, Asian countries failed to establish a joint monitoring and collaboration system to combat SARS because of a lack of information sharing service capabilities. Bilateral free trade treaties signed in Asia Pacific Region, gives more incentives for one government to provide services via internet to other governments to facilitate better information exchange. Governments which share resources could be more motivated to implement the inter-governmental G2G service (e.g. Singapore purchases water regularly from Malaysia).

4. CONCLUSION

Despite of the negative economical impact of the SARS outbreak, it has provided a big stage for e-government system to show its magic in Asian countries. It provided an excellent test ground to evaluate our current

development achievement and indicated a clear way to achieve the ultimate goal in e-government development. The successful operation of the e-government system during the SARS outbreak is a milestone in the history of e-government development. In some extent, the e-government system development could determine a government's effectiveness under emergency situation.

The e-government system started from some G2E and G2C applications, where the current focus has shifted to more valued-added G2B and G2G applications. The ultimate (mature) stage of the e-government development (See Figure 1) is to achieve a *Virtual Government* where all government service and interaction is available without any physical interaction. This is also coincidental to what the current ubiquitous network development tries to achieve.

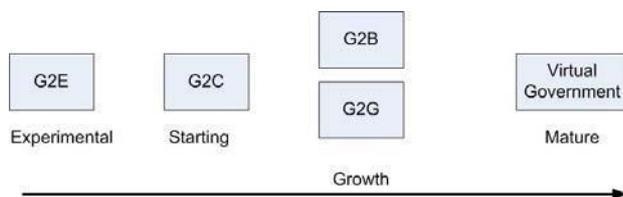


Figure 1: stages in e-government development

ACKNOWLEDGEMENT

The research is supported by various colleagues and students in the School Of Industrial Engineering, Georgia Institute of Technology.

REFERENCES

- [1] Eirinaki, M., & Vazirgiannis, M., "Web Mining for Web Personalization", *ACM Transactions on Internet Technology (TOIT)*, 3:1 Feb 2003, pp 1-27.
- [2] Fernandes, D.; Gorr, W. and Krishnan, R. "ServiceNet: One-Stop E-Government Services." *Proceedings of Seventh Americas Conference on Information Systems*, 2001.
- [3] Heeks, R. "Understanding E-Governance for Development", *Institute for Development Policy and Management, University of Manchester, I-Government Working Paper Series* No. 11, 2001.
- [4] Kenyon, J., and Vakola, M., "Customer Relationship Management: A viable strategy for the retail industry?", *International Journal of Organization Theory and Behavior*, 2003, 6:3 pp 329
- [5] Layne, K. and Lee, J. (2001). Developing fully functional E-Government: A four stage model. *Government Information Quarterly*, Vol.18, 2001, pp. 122-136.
- [6] Moon, M. J., "The Evolution of E-Government among Municipalities: Rhetoric or Reality?" *Public Administration Review*, Vol.62, No.4, 2002, pp. 424-433.
- [7] Osborne, D. and Gaebler, T.. *Reinventing government: how the entrepreneurial spirit is transforming the public sector*. New York: Plume.
- [8] Peppard, J., "Customer Relationship Management (CRM) In Financial Services", *European Management Journal* (18:3), 2000, pp. 312-327
- [9] Watson, R.T. and Mundy, B. A Strategic Perspective of Electronic Democracy. *Commun, ACM* 44, 1 (Jan. 2001), 27-30.

Analysis of Price Dispersion in E-Market: A Case Study

Zhiyuan Ge¹, Yong Shao²

¹ School of Economics and Management, Beijing University of Technology, Beijing 100022, China

² School of Software Engineering, Beijing University of Technology, Beijing 100022, China
{gezhy, mazel}@bjut.edu.cn

ABSTRACT

Some studies show that Internet markets are more efficient than conventional markets with respect to price levels, menu costs, and price elasticity, except substantial and persistent dispersion in prices on the Internet. Is price dispersion in Internet markets higher than that of conventional markets? This paper examines the nature of price competition in books among five online book-sellers in China, i.e. dangdang.com, joyo.com, bjbb.com, china-pub.com and welan.com. Unlike the researches before, we not only consider the products' prices, but also the whole price level. According to the investigations and analyses of 50 books price observations collected in August 2004, we conclude that the whole price dispersion in Internet markets is not higher than conventional markets, and high price dispersion on one sample is the results of firm's discrimination strategy. When the competition is very complete, the prices show a tendency to be consistent.

Keywords: E-Commerce, Price Dispersion, Internet Markets, Market Efficiency

1. INTRODUCTION

Electronic commerce is a significant force in the consumer economy. Consumers can now go online and comparison shop between hundreds of vendors with much less effort than in the physical world. One widely-reported prediction with respect to electronic commerce is that the availability of low-cost information on price — specifically the rise of comparison shopping agents — will lead all Internet retailers to charge the same price for mass-produced physical goods and that price will be approximately cost.^[1]

However, several studies have found that Internet markets are more efficient than conventional markets with respect to price levels, menu costs, and price elasticity, except substantial and persistent dispersion in prices on the Internet. This price dispersion was not consistent with Internet markets efficiency.

Is price dispersion in Internet markets truly higher than conventional markets? This paper approaches this question by investigating the price dispersion in Chinese electronic markets with examining the nature of price competition in books among five online book-sellers, i.e. dangdang.com, joyo.com, bjbb.com, china-pub.com and welan.com. In section 2 we discuss the researches of price dispersion in electronic markets. In section 3 we discuss the investigation in Chinese online book markets, and then in section 4 we compare it with traditional retail markets. In section 5 we discuss the phenomena of large price dispersion in online book

market and its competition, and then we present the basic results. In section 6 we conclude.

2. PRICE LEVEL AND PRICE DISPERSION IN ELECTRONIC MARKETS

An efficient market means in which consumers are susceptible to the changes of price, and product prices are susceptible to the changes of consumer demands, and operation costs of enterprise, product prices, marginal production costs and average production costs are equal. According this definition, the market efficiency can be measured by four parameters related to price: price level, price dispersion, menu cost, and price elasticity.

Generally, when consumers search costs decrease, product prices will decline and close to marginal cost. The less search costs, the lower price level, the less price dispersion, the higher market efficiency. According to the traditional economic view, by using search engine and electronic agent technology, consumers can search information of products more easily on the Internet, and compare the prices among many producers. Therefore, we would expect more violent competition and lower prices than in conventional channels. In addition, Internet is open to each enterprise, including big company and small firm, so it reduces enterprises' enter costs, and make the competition more intense.

Despite this presumption of increased competition, however, existing empirical works on the Internet have not been so supportive of the theory as one might expect. Because demand price elasticity and menu cost cannot be measured directly, existing work has mainly entailed collecting prices on and offline for a specific category such as books, and compared their price dispersion rate. Although the results from the works have not

conformed to the traditional views of reducing search costs. These studies have generally found large dispersion of prices online and prices either modestly lower or actually higher than their offline counterparts.^[2]

Early works yield conflicting results. Lee (1997) found that prices for second hand cars were higher on the Internet than in conventional channels. Brynjolfsson and Smith (2000) focused their research on the index of price level, price dispersion, and variety rate (menu cost) in online books and compact disks market for 1998-99. They found that Internet retailers had lower prices, smaller price adjustments, and larger or smaller price dispersion than conventional channels, depending on whether prices were weighted by proxies for market share. And their studies show that Internet retailer prices differ by an average of 33% for books and 25% for CDs, and some even reached 50%. Another study of interest is Clemons, Hitt, and Hann (1998) on the online travel industry. They found that online ticket agents engaged in significant product differentiation, i.e., agents responded to identical requests with different time/price pairs. Clay (2001) found that online bookstore had lower prices or the same, but had larger price dispersion.

Zhu^[3] proposed his interpretation frame on this paradox. He attributed the large price dispersion in E-markets to the weak position of consumer contrasted with the producer on deciding exchange prices. Although consumer can collect the information of retail price and related product prices, and have the decreasing tendency on prices, Zhu thought that E-markets reduce the search cost of online retailers at the same time. In his opinion, with lower search cost the online retailers can make a different price by customization and other strategies, to strengthen their market position and control consumer.

However, Zhu's theory lack evidences and is not appropriate in interpreting price dispersion, because on condition of complete competition, product prices of different companies tend to be consistent. Especially with improvement in their capability of obtaining information, they can respond to the market quickly. Zhu also did not explain how online environment improves enterprises' decision abilities, and enhances their price discrimination abilities.

3. INVESTIGATION OF PRICE DISPERSION IN ONLINE BOOKSTORES

In this paper, we will present the first empirical evidence on the impact of Internet competition on prices and dispersion offline. In this sense, our results are similar to the existing empirical works on search. By analyzing the data collected from five online bookstores, we are able to document the price level and price dispersion in Chinese electronic market. Moreover, unlike the researches before, we not only consider the products' prices, but also compare the whole price level.

We examine online bookstores because of the following three aspects: First, each book has fixed publishing price on it which we can see, though different retailers can sell with different prices. Second, Internet book retailers were the forerunner in electronic commerce, so the competition in this market is more complete than others. Third, other Internet retailing is not developed at present in Chinese.

Since Internet bubble had broken down in 2000, many online retailers disappeared. Nowadays, Internet book retailers have formed a big industry in China. The top 5 in total sale and sale categories are approximately Dangdang, Joyo, China-pub, Welan and Bertelsmann. There are many concrete bookstores engaged in online business, such as Shanghai book-city, Beijing book building. And some integration Internet retailers sale book, for example 8848.com.

We investigated the top four Internet book retailers and a mouse&brick bookstore: Dangdang, Joyo, China-pub, welan, and BJBB (Beijing Book Building) Online. Dangdang (www.dangdang.com) is one of the biggest and earliest online bookstores in China, which claims to be the biggest Chinese Internet mall in the world. Besides books Dangdang sells CD, VCD, DVD, software, magazines and general merchandises, and deals in C2C business. Nearly founded as early as Dangdang, Joyo (www.joyo.com) claims to be the biggest book and audio&video products mall on the Internet. Their business is similar to Dangdang, including books, music, video, etc., with the functions of community and personal auction. China-pub (www.china-pub.com) claims to be the most professional online bookstore in China, and mainly sells computer, communication, economics, management and foreign language books. Electronic books are its particular business. As a professional online bookstore Welan (www.welan.com) only sells books, and claims to be the most professional academic bookstore in China. Its primary business is restricted in Beijing. Bjjb Online (www.bjjb.com) is an Internet book retailer attached to BJBB.

We collected data from five companies mentioned above because of the following points. ① All of them except Welan are well known in China. ② All of their headquarters are in Beijing, and they have oriented the same client group in the same place. ③ All of them are professional in book retailing, and most of them sell more than 300 thousand kinds of books. ④ Except China-pub, others were founded at the end of 90s, and underwent the shock of Internet bubble break in 2000.

A data set of 50 popular books' prices collected in August 2004 in different fields was selected as samples to analyzing price dispersion phenomena. We chose popular books because most of them can be found in five online bookstores. Because of different operation strategies among five companies, if we select

professional books as samples, we cannot find them in all of bookstores, as a result, many data cannot be collected, and the data set will not be integrity, therefore we cannot analyze it.

Our sample books consist of three categories: economics, English learning and computer programming. One of rules to select book is popular we have mentioned before. Another rule is that selected book's prices distribute dispersal. We lean to select high price book, because consumers show different action in buying different price books. People do not hesitate in buying low price books even no discount. But if there is a large discount for higher price books, they maybe get a great customer value. Therefore, the average price of sample books is 48.2RMB.

In fact, because each bookstore differs from operation strategies, customer groups and professional fields, it is difficult to find fifty books which all five bookstores have and satisfy our standards. We had met many cases that cannot find in five bookstores at the same time when collecting data, for example, there is not the same edition of *Three Empire*, a famous classic novel came down from Ming dynasty.

Although Joyo claims to be the biggest book and audio&video products mall on the Internet, we find its book categories are fewest in five companies. Dangdang faces to the masses. China-pub faces to professionals in computer field, economics and management. Welan was founded by a few students of Tsinghua University in 1999, and mainly faces to students and teachers of university in Beijing.

Table 1 shows some index we obtained.

Table 1. Price comparison of five online bookstores

Index	Value
Max Discount (Off)	52%
Min Discount (Off)	0%
Average Discount (Off)	20%
Max Price Dispersion	46.2%
Min Price Dispersion	10%
Average Price Dispersion	15.2%

Where price dispersion p equals

$$p = \frac{\max\{i_t\} - \min\{i_t\}}{\max\{i_t\}} \times 100\% \quad (1)$$

and average price dispersion p_a equals

$$p_a = \frac{\max\{\sum i_t\} - \min\{\sum i_t\}}{\max\{\sum i_t\}} \times 100\% \quad (2)$$

We analyzed the data and worked out the following results:

- ò Big discount is given in some books. The biggest discount is 48%, and the average discount is 80%. It possibly relates with the samples we selected,

because some books, especially computer books, are expensive, and have a big discount space. Some books are sold with very low price because of sales promotion.

- ò There is high price dispersion in online book markets. Max price dispersion reaches 46.2%, and min price dispersion is 10%, and average price dispersion is 15.2%. The max price dispersion is possibly due to sales promotion. Our results showed difference with Brynjolfsson's studies^[5] which said average price dispersion was 33%.
- ò Compared with conventional bookstores, prices were lower in online bookstores. As a conventional bookstore, BJBB spreads their brand with starting online business and changed to a click-and mortar company. Other four bookstores are virtual company, and their average selling price is a little low.

However, after analyzing the data set we collected further, we find that the average price dispersion is only 6.8% if we remove the data of BJBB! This is an acceptable result which means the price dispersion of online market is not higher than conventional market.

4. FARTHER STUDIES: COMPETITION BETWEEN ONLINE BOOKSTORES AND OFFLINE BOOKSTORES

Some of the results above are consistent with researches before, such as there is large price dispersion of one product, and max price dispersion reaches more than 46%. Compared to 90% discount in conventional bookstores (except for special bookstore), online bookstores have a large space in making a price. All those data seem to support the theory that electronic market has not been exhibited more efficient than conventional market, though it would be.

Because there is not a reasonable interpretation as yet, we have to carry on our research in another method. We change our research focus on the competition between online bookstores and offline bookstores. We investigated five offline bookstores which are located at and near Tsinghua University. We selected them basing on the following two aspects. On one hand they open to the same client group — students and teachers on the campus and at other universities hereabout, so they were intensely in competition with Welan. On the other hand, we are familiar with the evolvement of competition between them.

The main promotion measure of online bookstore is discounting. Welan resorted to it to absorb clients when it was founded. At the beginning (1999) Welan sold English-studying books published by New Oriented School (those reference books have been very popular in Chinese students who want to pass the English

examinations such as TOEFL, GRE, IELTS, etc.) at a 79% discount, and it advertised widely. Soon (about 3~4 months after) the five conventional bookstores adjusted the same books' prices from 90% discount to 80% discount, and then fell to 79% discount. In fact the competition spread to the bookstores further.

The competition mentioned above could be regarded as a game between online bookstores and offline bookstores. Online bookstores were a new entrant who had to introduce some new measures on purpose to absorb clients. So Welan chose discount. As a virtual shop, online bookstores need not a physical place, and they have all advantages that a virtual company have contrasted with offline ones.

When Welan opened the five conventional bookstores had not minded the new entrant. But after they found their turnover fell and had lost some clients gradually, they had to act. Firstly they do not think that they had to make the price as low as their new rival because of their advantages, for example, clients can obtain books immediately, so they just give 80% discount. But they found the clients who came to rivals were not easy to return, even they adopted lower prices measure then before, and it is inadequate to win the competition. Finally, the five conventional bookstores reduced their prices as the same as Welan.

The brick-and-mortar rivals' action made Welan lose its price advantage, so Welan adopted measures to keep customers. They sold books with 75% discount. Here online bookstore's profit was close to offline bookstores, and the customers' number of two kinds of bookstore tended to be steady gradually. Therefore, the competition was in equilibrium.

The game process was shown in table 2 and figure 1.

Table 2. Payment matrix of online and offline bookstore

		Conventional bookstore	
		Falling price	Keeping
Internet bookstore	Falling price	1,1	2,0.5
	Keeping	0,1.5	0,3

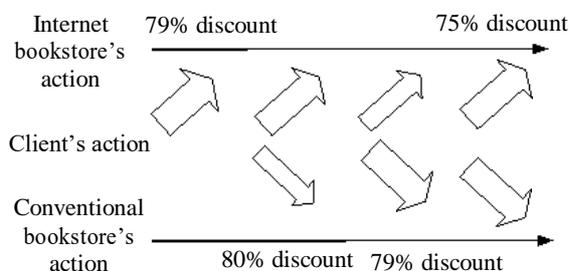


Figure 1. Competition between online and offline bookstore

From the investigation we conclude:

Under the competition in which client group overlaps and information transfers sufficiently, no matter what kind of bookstore, the price dispersion is low and prices will tend toward the same. It is consistent with the theory that Internet will improve market efficiency. Consumer strength of bargaining is improved instead of declined as some researchers said. Decision capacities of enterprise are enhanced too, but not cost consumer's bargaining strength.

5. PRICE DISPERSION, COMPETITION IN ONLINE BOOKSTORES

With further analysis, we find that on condition sufficient competition price will trend to the sameness quickly. Our results differ from the former researchers' conclusion that the price dispersion of online book market is larger than offline. But our results based on different backgrounds of bookstores: we investigate the bookstores which have the same client group — they are similar in consumed level, knowledge level, reading taste, etc.

Therefore, we conclude that why there is high price dispersion that former researchers found due to the following points:

- ò Different Internet bookstore can carry on their price discrimination because of different client groups.
- ò Because of cost advantage Internet bookstore have a large spare in making a price, so there is big price dispersion.

Few consumers only buy one book once. For those consumers whatever bookstore is the same, and they tend to choose the cheapest one. However, for those consumers who need more than three books in several categories, it is difficult to choose bookstore (only consider price factor) because different bookstore have different promotion strategies. For example, Dangdang resort to using a very low price in best seller as a method to attract consumers, and as a professional bookstore the discrimination strategy of Welan is giving big discount for all. When books a consumer wants to buy include some best sellers and catena, he will find some of them should buy from Dangdang but others should from Welan, and if he selects the two it will take him more time and more delivery cost. It is very difficult to decide which bookstore he prefers.

So we attribute the price dispersion of online book markets to different client groups (purchasing power, knowledge level, professional background, etc.) and the

relationship between different books (many people have a few kind of reading taste)

6. CONCLUSION

In this paper we have examined some Internet bookstores in Aug 2004 and reviewed the competition between *welan.com* and some conventional bookstores from 1999 to 2000. We find that the phenomena that price dispersion of Internet markets is higher than conventional markets only exist in one book or single product instead of integer, and due to the discrimination strategies of enterprise. On the contrary, the whole price dispersion is not high, and on condition of sufficient competition different companies' prices trend to be the same. The data also show, consistent with the theory, that the efficiency of online markets is higher than conventional markets.

Another study of the competition between a virtual company and some conventional shops from 1999 to 2000 indicates that competitors adjusted their prices following the virtual shop. The results reveal that Internet market enhances the competition because information can be gotten easily.

Certainly, our studies are insufficient. We have only investigated five typical Internet bookstores and selected samples purposefully. In fact, we should consider consumer purchasing power, motivation, and

other consumer's characteristic, as well as market development, and enterprise's relative business, etc., which will infect our results.

REFERENCES

- [1] Karen Clay, Ramayya Krishnan, etc., "Retail Strategies on the Web: Price and Non-price Competition in the Online Book Industry", *kclay@andrew.cmu.edu*.
- [2] Jeffrey R. Brown (Kennedy School of Government, Harvard University), Austan Goolsbee, "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry", 2000.
- [3] Zhu Tong, "Paradox of Internet Market Efficiency: A New Interpretation Frame", *World Economics*, No.2, 2002, P:12~20 (in Chinese)
- [4] David VanHoose, "E-commerce Economics", Thomson Learning, 2003
- [5] Erik Brynjolfsson & Michael D. Smith, "Frictionless Commerce? A Comparison of Internet and Conventional Retailers", <http://ecommerce.mit.edu/papers/friction>, 1999
- [6] Karen Clay, Ramayya Krishnan, and Eric Wolff, "Prices and Price Dispersion on the web: Evidence from the Online Book Industry", National Bureau of Economic Research Working Paper 827, May 2001
- [7] Han Hong, Matthew Shum, "Can Search Costs Rationalize Equilibrium Price Dispersion in Online Markets?", 2003.

Current State of China's Online Gaming Industry and the Obstacles in Development

Bin Qiu¹, Shaoqin Sun², Cunyan Fan³, Shujen Chen⁴

¹School of Economics and Management, Southeast University, Nanjing 210096, China
binqiu@163.net

²Dept. of Economics and Trade, Nanjing Information Engineering University, Nanjing 210048, China
shaoqinsun@yahoo.com.cn

³Netease Company, Guangzhou 510665, China
hivans@sina.com

⁴California State University, Fullerton, U.S.A.
Schen@exchange.fullerton.edu

ABSTRACT

This paper aims at identifying the current state of China's online gaming market, and exploring the obstacles that keep China from developing this industry further on the basis of analyzing the value chain of online gaming industry. In the meanwhile, a case study will be made about the domestic leading online gamer—Shanda corporation, a listed corporation on NASDAQ. And in the end some conclusions will be made on how to assess the online gaming industry in China and how to remove the obstacles in accelerating the development of online gaming industry in China.

Keywords: online gaming industry, current state, China, obstacles, Shanda Corporation

After the dot-coms gradually lost their wizards in the past few years, the new economy firms never stopped seeking the potential lucrative industry that can make them "recovered". Recently, they have found a booming industry that can lure people from other things—online games. According to a recent survey made by the International Data Corporation, or IDC, China's online gaming market grew 46 percent in 2003, and in the next two years its Massive Multiplayer Online Role-Playing Games (MMORPG) are expected to dominate its online gaming market.

1. MARKET SCALE OF CHINA'S ONLINE GAMING INDUSTRY

1.1 Global Online Gaming Market Size

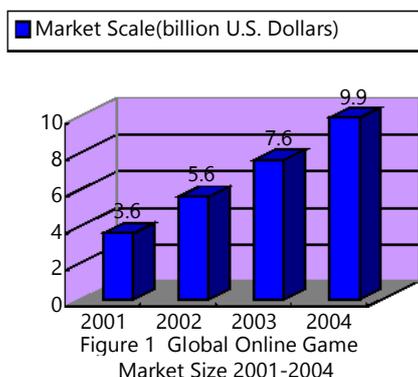
The increase in broadband connectivity to the home has an explosive impact on the world of online gaming, thus, the online gaming sector is growing tremendously as the availability of faster processing hardware enables the development of more advanced games.

According to the most recent statistics, the global online game market scale reached 7.6 billions U.S. dollars in the year of 2003, and by estimate, the total revenue of online gaming industry in 2004 will reach 9.9 billion U.S. dollars (see Figure 1).

1.2 Current State of China's Online Gaming Market

In 2003, China has become the second-largest country of installing PCs and has the third-largest internet users. According to the latest survey made by CNNIC in July 2004, the internet users in China has become 87 million, among whom 31.1 million people use broadband as the method of accessing internet. There is no doubt that high speed Internet accession is another reason why China is growing so rapidly in online gaming industry. [1]

Game developers in China are hoping that they can occupy more market share currently dominated by Korean companies. Legend of Knights, the first online game made by Chinese company has been quite successful with peak simultaneous players of 80,000. A variety of statistics made by iResearch Consulting Company recently demonstrate that internet games have become a great business in China. Earlier this year, of the Top 100 tycoons on the Forbes list, three from the IT industry confirmed that most of their revenue had come from the online gaming industry. [2]



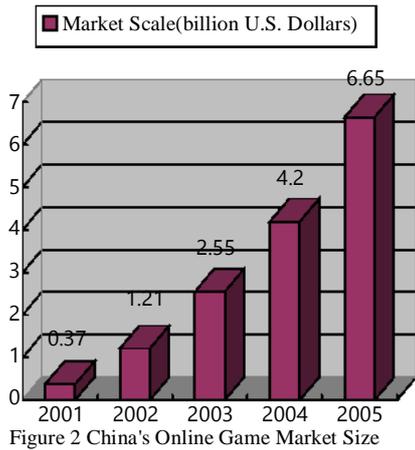


Figure 2 China's Online Game Market Size 2001-2005

Statistics show that the total revenue of China's online gaming industry reached 2.55 billion U.S. dollars in 2003. By estimate, the volume of total output of this industry in China will reach 6.65 billion U.S. dollars in 2005 (see Figure 2). Online gaming market estimate has also shown that China has the highest potential of growth in the Asia-Pacific region. The most optimistic remark on China's growth from IDC says that China will probably become the largest online gaming market in 2007 with

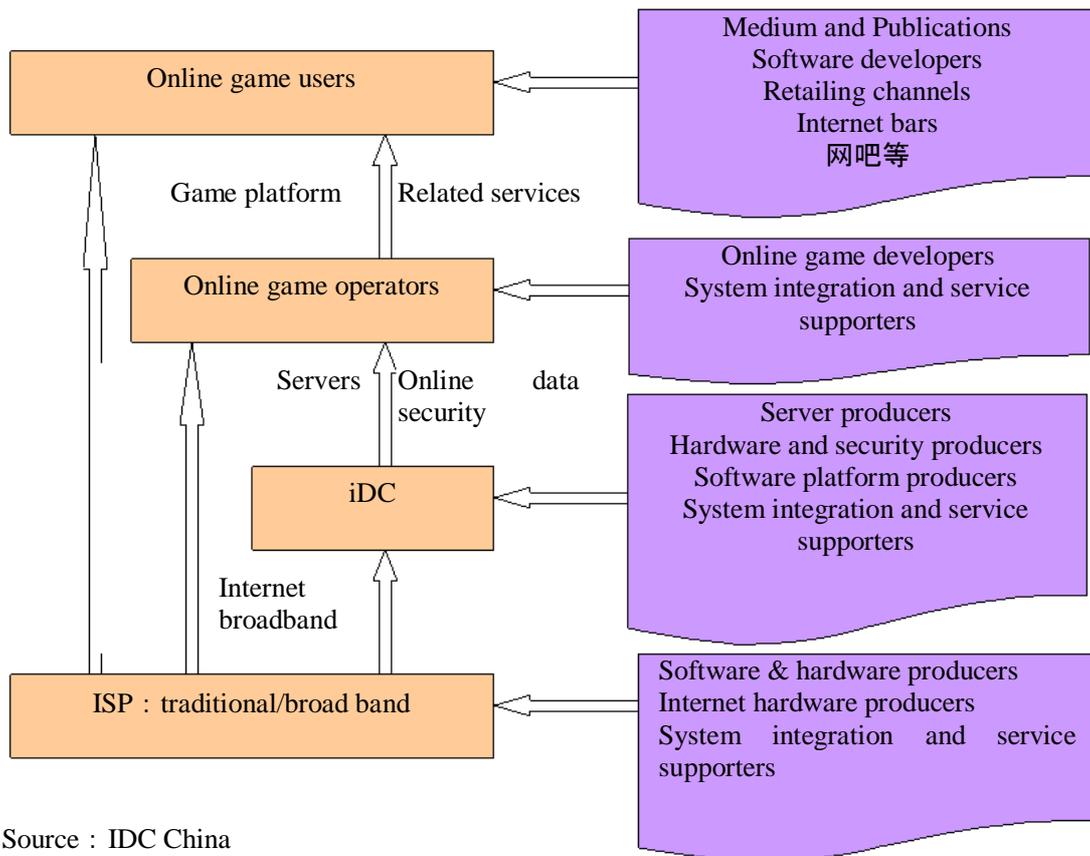
availability of broadband and wireless networks acting as the key to growth.

2. VALUE CHAIN OF ONLINE GAMING INDUSTRY AND THE POSOTION OF CHINESE FIRMS

As we know, the whole value chain of online games industry can be divided into the following categories: game developers, game operators, marketing channels, ISPs and peripheral services.

The main chain of the online game industry is as follows: online game players, online game operators, online game developers and ISPs; and the supporting chain of the online game industry is: internet Cafés, online game sellers, peripheral products for online games, mediums and publication industries, computer software industry, etc (see Figure 3).

In terms of the main value chain, the Chinese firms mainly concentrate on the marketing of online games, while the hi-end of the value chain, i.e. the development of online games is under the control of foreign companies especially at the hand of Korean companies.



Source : IDC China

Figure 3 Value Chain Of Online Gaming Industry

3. THE MAIN OBSTACLES THAT KEEP ONLINE GAMING INDUSTRY IN CHINA FROM DEVELOPING FURTHER

3.1 Main Obstacles from the Perspective of End-Users of Online Games

Although the internet infrastructure construction has made great progress in recent years, connect speed is still the first factor that influences end-users in choosing what games they play. The fact that 69.4% of the Chinese online game players think that connect speed is the main factor on their game choosing demonstrates that the broadband service in China is not so satisfactory. Cost is the second factor that affects online game players (39.3%). Besides, difficult level, story design and visual design are roughly equal with about 30-40% players deem them as factors that influence them on choosing games respectively (See figure 4).

3.2 Lack of Support from Government with Unclear Industry Policies

Apart from the negative attitude from the public who

tend to think that online gaming industry might be a potential threat that keep their children away from studying, the government also shows a vague attitude toward this industry. For example, it's not an easy thing to get licensing of setting up Internet Café in China. Another fact that can not be ignored is that the internet accession fee is still high compared to the income of the internet game player (Most of the game players are young students who still rely on their parents). There are still other problems such as piracy problem, monopoly problem in telecommunication industry, etc. [3]

3.3 Weak Domestic Online Gamers at the Lower-end of Value Chain

At present, the Chinese online gaming market is still dominated by Korean companies. The domestic online game developers are facing stiff competition from Korea, Japan and other countries. Most of the Chinese online game developers lack capital and comprehensive abilities of developing and operating in this industry. Some companies are very slow in renewing their games and the R&D activities are at a lower level compared to the above mentioned advanced countries.

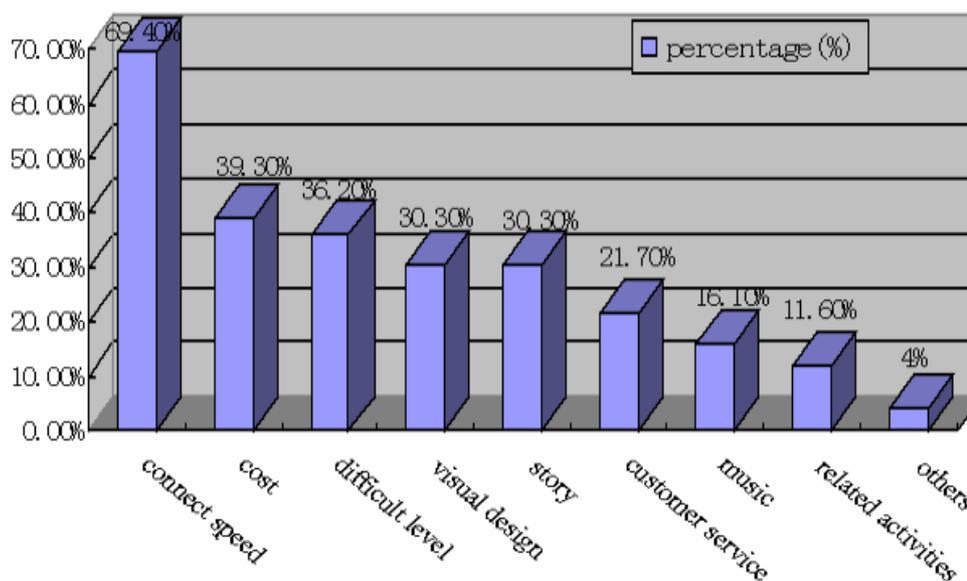


Figure 4 Factors That Influence Players On Choosing Games

4. CASE STUDY—SHANDA CORPORATION, AN OBSTACLE REMOVER IN CHINA'S ONLINE GAMING MARKET

Shanda is the largest operator of online games in China. This Corporation offers a portfolio of online games that are licensed from third parties and that are developed in-house. The games that Shanda offers include The Legend of Mir II, which is the most popular online game in China according to users surveyed by International Data Corporation ("IDC"), The World of Legend, which is Shanda's leading in-house developed game, and BNB, which was the highest ranked casual game in IDC's survey of the top ten most popular online games in China in 2003. Shanda is headquartered in Shanghai, China.

4.1 Shanda's IPO Strategy

In May 2004, Shanda Interactive Entertainment filed with US regulators for an initial public offering valued at up to 200 million US dollars. According to a preliminary prospectus filed with the U.S. Securities and Exchange Commission, Goldman Sachs underwrote most of the IPO for Shanghai-based Shanda. Other groups like Piper Jaffray and the Hong Kong and Shanghai Banking Corporation or HSBC also took part. Shanda said it has applied to list its American depository shares on the Nasdaq stock market under the trading symbol "SNDA".

The IPO of Shanda brought this company both fame and wealth. And this had made Shanda a pioneer among Chinese online gamers. Also, Tianqiao Chen, the CEO of

Shanda Corporation has become the richest person in China according to the shares he holds and the most recent market price of SNDA on NASDAQ.

4.2 Shanda's M&A Strategy

In August, 2004, Shanda acquired Hangzhou Bianfeng Software Technology Co., Ltd. Bianfeng is a leading developer and operator of chess and board games in China. Currently, Bianfeng offers over 50 different games, which attracted over 200, 000 peak concurrent users in July 2004.

The acquisition is based on the fact that board games are deep rooted in Chinese culture and have a large and loyal users especially among mature age groups. The acquisition of Bianfeng adds another important segment to Shanda's casual games platform and greatly broadens Shanda's user base demographics. [4]

4.3 Shanda's Marketing Strategy

The focus of Shanda's marketing strategy is the so-called "experience marketing" style. This style means that Shanda Corporation attaches great importance in finding their customers' feelings and preferences in playing their online games. For example, the data base naturally generated from the records in the process of consumers' playing can be of great help in analyzing what kinds of games they like to play, at what price level customers can afford, and at which time periods during a day the customers frequently log in, etc. The above mentioned information can help Shanda Corporation do better job to improve their services to seize a clearer group of target market.

Another factor that leads to Shanda's success in its marketing strategy is that this company uses internet Cafés as their marketing channels. By using internet Cafés which are very popular in China, it's relatively easy for Shanda to establish its marketing channel in a short period of time. In the meantime, Shanda installs E-Sales system in each internet Café to build up an integrated information system among the distributors, and this helps a lot in doing business statistics and thus giving different discounts to distributors according to their business scale. Because there is no logistics problem in online gaming industry, Shanda utilizes banks and post-offices as its channel of collecting money for playing games.

5. CONCLUSIONS

The online gaming industry in China is still at an early stage, therefore, the online gaming market in China is immature at present. The main reason lies in the fact that the upstream of this industry is still dominated by foreign firms.

The online gaming industry has a long value chain including many industries such as medium and publication industry, telecommunication industry, hardware and software industry, retailing channels, etc, and the market opportunity is tremendously huge.

The main obstacles of online gaming industry in China include internet accession infrastructure to be improved, unclear industry policies and high industrial barriers and the internal reasons of Chinese online gamers including lack of capital, weak R&D abilities and relatively small scale, etc.

Some domestic online gamers in China have grown more and more strong. The Shanghai-based Shanda Corporation has set a good example in its IPO strategy, M&A strategy as well as its marketing strategy. We believe that more and more domestic corporations like Shanda will achieve great success in China's online gaming industry in not a long future.

ACKNOWLEDGEMENT

ICEB2004 is supported by the National Natural Science Foundation of China.

This paper is subsidized by "Excellent Young Teachers Subsidy Project" of Southeast University and the School of Economics and Management, Southeast University

REFERENCES

- [1] CNNIC, *14th Statistical Survey Report on the Internet Development in China*, July 2004
- [2] iResearch, *China Online Game Research Report*, http://iresearch.com.cn/online_game/
- [3] Wu Lianfeng, *A Research on China's Online Gaming Industry*, www.iec.com
- [4] Shanda Interactive Entertainment Limited, *Shanda Acquires Casual Game Operator Bianfeng*, http://biz.yahoo.com/prnews/040802/cnm004_1.html

Developing a B2B E-Commerce Evolution Model -- The Case Studies of Seven Industries in Taiwan

Vincent Ming¹, Guohong Chang², Chuehju Wu³, Chengying Lin⁴

¹ Department of Information Management, National Kaohsiung University of Applied Sciences,
Taiwan, China

² Information Service Industry Association, Taiwan, China

³ Graduate Institute of Commerce, National Kaohsiung University of Applied Sciences, Taiwan, China

⁴ E-Commerce Research Center, National Kaohsiung University of Applied Sciences, Taiwan, China
{vincming, a1092341116, zfoffice01}@cc.kuas.edu.tw, cisa@mail.cisnet.org.tw

ABSTRACT

The e-commerce environment is changing fast and getting mature, the internet technology enabled evolution created many new business organizations and new business models. It also caused new competitions and new management challenges. Although B2B e-marketplace have been promoted as tools for reducing transaction costs, streamlining process efficiency, and enabling virtual collaboration among partners, there are still not many successful B2B e-marketplaces in Asian. This research tried to develop an evolution model based on the IOS (Inter-Organizational Systems) model defined by Benjamin (1990)(1) and Applegate et al, (1996)(2). Through over 100 firms' case studies in seven industries of Taiwan, the patterns of business partnership structures and the path of e-marketplace evolution were found. In order to explore the factors effected e-marketplace evolution, this paper summarized the possible impact factors based on 'Transaction cost theory', 'Resource dependent theory', 'Institution theory', and the 'Contextual Influences' which were highlighted as three levels in previous study – socio-cultural, national/regional, and structural (Hsiao and Ming, 2002)(3). The questionnaires were developed and mailed to 980 firms on the list of class-A members of 'Taiwan Taiwan Electric and Electronic Manufacturers' Association.' 178 valid questionnaires were received and investigated. The data analysis result suggested that factors of institutionalization effect may be one of the most significant impact dimensions of e-marketplace evolution.

Keywords: Inter-Organizational System, e-marketplace, Industry Value Chain

1. INTRODUCTION

The e-commerce environment is changing fast and getting mature, the internet technology enabled evolution created many new business organizations and new business models. Beside a mass of merit and chance, internet also created new competitions and new management challenges. The enterprises could connect to supplier, business partners and customers through internet. This could not only increase the efficiency and reduce the cost of business transactions but also create the changes to the value chain of enterprise and the supply chain crossed different organizations. There has been a rising interest in using e-marketplaces (electronic marketplaces) to enable closer B2B (Business-to-Business) collaboration (Hsiao and Ming, 2002)(3). Many IT solution providers and government agents put lots of effort and investment to encourage firms to implement e-business especially the B2B e-marketplace. But market analysts have reported that the majority of the e-marketplaces have either closed down or are facing difficulties in attracting buyers and suppliers (Markus and Soh, 2002)(4). The study is an initial attempt to building a three dimension evolution model to investigate the evolution path of the B2B e-commerce development progress and the reasons for the success or failure.

2. THE IOS REVOLUTION

The popularity of IOS (Inter-Organizational Systems) started in the late 1960s. Felix Kaufman (1966)(5) first described the "information partnerships" through the information systems cross the Organizational Boundaries. Barrett and Konsynski (1982)(6) highlighted the value of IOS is to share information cross the organizations. Cash and Konsynski (1985) (7) defined IOS as: 'the automatic information systems shared by more than two companies.' They stressed that IOS can improve the productivity, flexibility and competition advantages. Malone, et al., (1987)(8) classified the IOS as electronic hierarchy and electronic market. The electronic hierarchy is for the control, communication and coordination between the different hierarchical levels of organizations. The electronic market is to support the market functions between different organizations. Bakos (1991)(9) based on the function classified the IOS as 'Information Links' and 'Electronic Market.' The information links are the 'Communication Channels' between the organizations in the industry value chain. The electronic market is the 'Intermediaries' or 'Marketplace' for buyer and seller.

Benjamin (1990)(1) divided the IOS into the two dimension matrix as figure 1. Benjamin claimed that the IOS implementation of the organization is an evolution process, the type of IOS will be changed from time to

time as the business changed. Konsynski (1993)(10) based on the 'Interaction Patterns'to divide the IOS into one-to-one, one-to-many, and many-to-many. Applegate, McFarlan and McKenney (1996) in their book--*Corporate Information Systems*, created a two dimension model to describe the IOS evolution path from single sales channel to electronic marketplace. Figure 2 list these two dimensions as 'Interaction Patterns'and 'Levels of Control.'

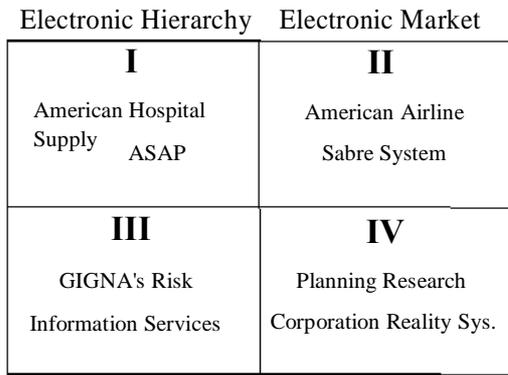


Figure 1. IOS evolution (adopted from Benjamin, 1990)

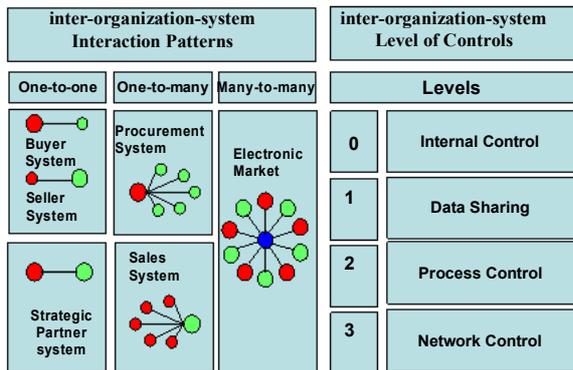


Figure 2. IOS Interaction Patterns and Level of Controls (Adopted from Applegate et al, 1996)

3. DEVELOPING A B2B E-COMMERCE EVOLUTION MODEL

This paper try to develop an evolution model by integrated the frame work created by Benjamin and Applegate et al., (Figure 3). The cases of America Airline (AA) and America Hospital Supplier Company (AHSC) were marked in the model to illustrate the path of 'evolution.'Both the AA and AHSC cases showed the IOS evolution started from Data sharing/Transaction processing to Network Control/Collaboration in the Control Level, and from one-to-one/electronic hierarchy to many-to-many/ e-marketplace.

This evolution model were tested and confirmed by over 100 cases study in seven different industries including grocery (30 companies), costume and accessories (8

companies), electronic (20 companies), books (14 companies), pharmaceuticals and cosmetics(12 companies), logistics(10 companies) and tourist(7 companies) industries. With the support from the Information Service Industry Association of Taiwan, a team of researchers and practitioners investigated the comprehensive files of over 100 firms that implement their B2B/IOS with their business partners during the period of three years from 2002 to 2004. The results confirmed the development of IOS in different industry do follow the similar path as AA and AHSC. The cases of Electronic, Books and Grocery are presented in Figure 4.

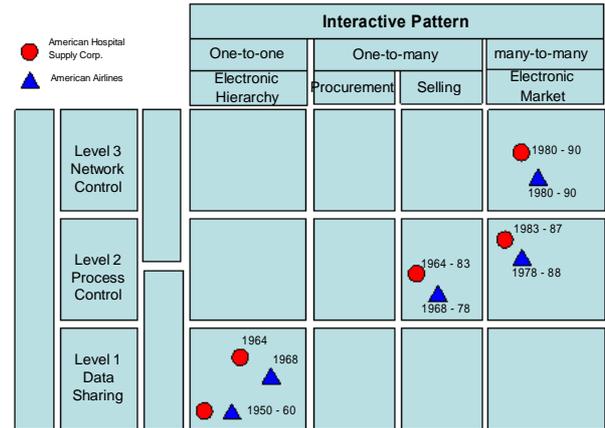


Figure 3. IOS evolution progress (the cases of America Airline and AHSC)

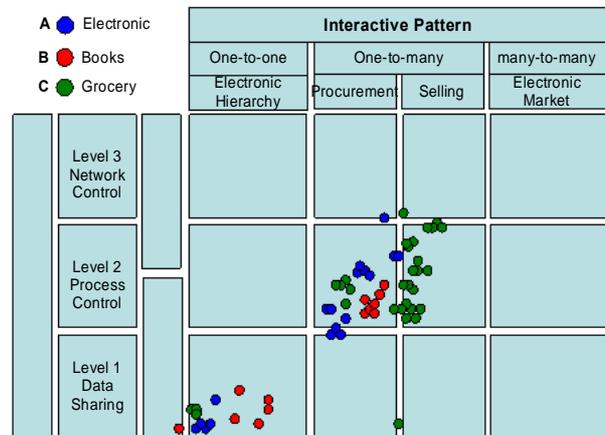


Figure 4. IOS evolution progress (the samples of three industries in Taiwan)

Although the progresses of IOS/B2B system development are different between different companies, there seems to be some similarities in the same industry. The overall maturity of e-business development in electronic industry and grocery industry are better than the books in average. In order to understand the reasons of these similarities, the researchers need to further explore the business partnership structures and the readiness of B2B e-commerce in different industries. According to the documents provided by the Information Service Industry Association of Taiwan, the researchers can investigate the important connections

between B2B e-commerce development and industry value chains. Over 100 copies of business plan and system implementation documents were reviewed and categorized. Five patents of business partnership structures and the path of e-marketplace evolution were found. Those five business partnership patterns— which were named as central competition, cross competition, intermediary dependant, crisscross and proprietary (figure 5); reflect the power structures in the supply chain of different industries. How are these different power structures effect the B2B e-commerce development in different industries? Lots of academic research mentioned many factors that might impact the success of B2B e-commerce. Could we find out the connections between those influence factors and the B2B e-commerce evolution?

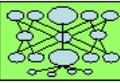
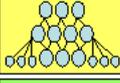
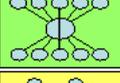
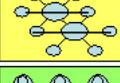
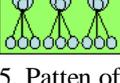
Type	Structure	Industry	Characteristics
Central Competition		Food? Grocery Retailing	Fewer leading companies with competition, each Competitor owns strong power in the Supply chain.
Cross Competition		3C (Electronics)	Many leading companies, each Competitor owns their channels and no one can control the supply chain.
Intermediary Dependant		Books	No leading companies in the retailing channels and suppliers Intermediaries owns the power to integrate the supply chain
Crisscross		Travel/tourist Logistics	Most of the leading companies are competitor and partner to each other. It's very difficult to integrate the supply chain.
Proprietary		Cosmetic Pharmaceuticals Costume	The branded leading companies Control the market and own their proprietary channels,

Figure 5. Patten of Industry Business Partnership

4. EFFECT FACTORS OF THE EVOLUTION

In order to explore the factors effected e-marketplace evolution, this paper summarized the possible impact factors based on ‘Transaction cost theory’, ‘Resource dependent theory’, ‘Institution theory’, and the ‘Contextual Influences’ which were highlighted as three levels in previous study – socio-cultural, national/regional, and structural (Hsiao and Ming, 2002) (3).

All the factors and their related articles and the theory based are listed in table 1.

Table 1 Factors that Effect IOS/B2B e-commerce

Theory	Reference
Transaction Cost Theory	
Transaction cost	Dietrich (1994) (11) Mahoney (1992) (12)
Uncertainty	Robertson and Gatignon (1998) (13)
Dependence	Barney(1990) (14) Williamson(1988) (15) Anderson and Narus (1990) (16) Zaheer and Venkatraman

	(1994) (17)
Resource Dependent Theory	
Dependent	Emerson(1962) (18) Pferffer&Salancik(1978) (19)
Asset Specificity	Sengupta, Krapfel and Pusateri (1997) (20) Zaheer andVenkatraman (1994)(17)
Institution Theory	
Coercive	DiMaggio&Powell(1983) (21) Forgarty(1992) (22)
Mimetic	DiMaggio&Powell(1983) (21) Forgarty(1992)(22)? Galashiewicz (1985)(23)
Normative	DiMaggio &Powell(1983) (21) Lachman & Aranya (1986) (24)
Contextual Influences	
National/Regional	Avgerou (2001) (25) Hsiao and Ming (2002) (3)
Socio-cultural	Kumar(1998) (26) Hsiao and Ming (2002) (3)
Structural	Markus & Soh(2002) (4) Hsiao and Ming (2002) (3)

5. DATA COLLECTION AND ANALYSIS

Based on the literature review and the summarized factors listed in table1, the questionnaires were developed to find out the possible factors effect the B2B e-commerce development in Taiwan. Due to the cost and time constrains, the survey is just for the electronic industry in the first stage. Data were collected from the list of class-A members of ‘Taiwan Taiwan Electric and Electronic Manufacturers’ Association.’and mailed to 980 firms between February 2004 and April 2004. There are 178 valid questionnaires were received and investigated.

Simple bi-variable associations among the collected variables reported as Spearman rank-order correlation coefficients. Factor analysis was used to describe the interrelationships of multiple variables. This kind of analysis is a multivariable analytic technique that users variable(with unknown correlations) to create a new set of variables called ‘factors’(Sharma 1996) (27).

Examining the correlation among the studied factors revealed that there was significant overlap among various subgroups of factors. Next, with factor analysis it was possible to investigate the number of various subgroups and to identify what these subgroups represent conceptually.

There are fifty-four items in questionnaires and high

correlations among many of these items were observed, suggesting the appropriateness of data reduction techniques. To reduce variables, we began our investigation with the use of principal components analysis followed by varimax rotation for the thirteen indicators.

On ‘Transaction cost theory’, the factor analysis provided four factors with eigenvalues than 1.00(4.46, 2.11, 1.38, 1.10) . These four factors explained 64.64% of total variance. On ‘Resource dependent theory’, the factor analysis provided four factors with eigenvalues than 1.00(4.94, 2.08 ,1.27) . These three factors explained 55.23% of total variance. On ‘Institution theory’, the factor analysis provided two factors with eigenvalues than 1.00(3.56, 1.13) . These two factors explained 66.92% of total variance. On ‘Contextual Influences’, the factor analysis provided three factors with eigenvalues than 1.00(4.17, 1.95, 1.56) . These three factors explained 64.08% of total variance. On ‘Electronic Integration performance’,the factor provide one factors with eigenvalues than 1.00(5.05) . This factor explained 63.15% of total variance.

Table 2 lists the alpha coefficients for the factor as well as the predicted alpha coefficients except ‘Technological replacement’. Therefore, the below results suggest that the ‘Technological replacement’ factor that low homogeneous were deleted. On this background, it seemed appropriate to extract twelve factors in the present study.

Table 2 Alpha Coefficients

Factor	Alpha	Factor items
Transaction Cost Theory		
Dependence	0.782	5
Technological replacement	0.375	5
Transaction cost	0.778	3
Uncertainty	N/A	1
Resource Dependent Theory		
Resource dependent	0.794	6
Asset Specificity	0.806	7
Business Dependent	0.585	2
Institution theory		
Normative& Mimetic	0.791	4
Coercive	0.813	3
Contextual Influences		
National/regional	0.861	6
Socio-cultural	0.820	3
Structural	0.696	3
Electronic Integration Performance		
Integration performance	0.916	8

A further application of factor analysis uses the factors in a regression analysis. The regression coefficients obtained hereby would be more stable than the regression coefficients in the original variables. The eleven factors given above were independent in the

regression model and Electronic Integration performance was dependent. Eleven factors explained 51.8% of the variability in data. Resource dependent, Business Dependent, Normative and Mimetic, Coercive, national and structural were significantly. The results are shown in Table 3.

Table 3 Regression Analysis

Factors	Coefficient	Prob.
Dependence	2.711E-02	.638
Transaction cost	3.154E-02	.616
Uncertainty	3.972E-02	.488
Resource dependent	.253	.000
Asset specificity	-9.351E-02	.187
Business Dependent	.159	.011
Normative & Mimetic	.305	.000
Coercive	.146	.038
National	.193	.005
Socio cultural	-5.418E-03	.928
Structural	-.218	.001

6. CONCLUSIONS

The data analysis result suggested that factors of institutionalization effect may be one of the most significant impact dimensions of B2B e-marketplace evolution. This result explains the difference of e-commerce development in seven industries. The stronger of the power structure in business partnership, the easier for the leading companies to integrate the industry value chain. Coercive, Mimetic and Normative of institutionalization effect will make the firms of the same industry reach the similar status in the B2B e-commerce.

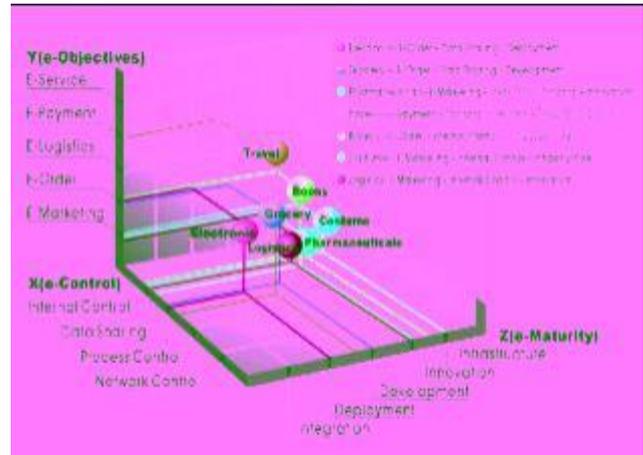


Figure 6. B2B e-Commerce Evolution Progress (The Samples of Seven Industries in Taiwan

The evolution model was modified to reflect the environment factors that might impact the adoption of B2B e-commerce. The final three dimensions model indicates that the objectives, the control levels and the maturity of the B2B e-commerce will differentiate the status of B2B e-commerce in different companies and different industries. The summary of the seven

industries position in the e-marketplace evolution is presented in figure 6.

ACKNOWLEDGEMENT

This research is supported by the National Science Council (Project No. 92-2416-H-151-004) and Information Service Industry Association, Taiwan.

REFERENCES

- [1] Benjamin, R. I., de Long, D. W. and Scott Morton, Ms S., "Electronic Data Interchange: How much Competitive Advantage?", *Long Range Planning*, Vol.23, No.1, pp29-40, 1990.
- [2] Applegate, McFarlan and McKenney, "Corporate Information Systems Management : Text and Cases ", Fourth Edition, pp156-186, 1996.
- [3] Ruylin Hsiao and Vincent Ming, "The Asian Difference in B2B E-Marketplace", *Proceedings of the Second International Conference on Electronic Business*, Taipei, pp532-534 (full paper in CD-ROM), 2002.
- [4] Markus, M. L. and Christina Soh., "Structural Influences on Global E-Commerce Activity", *Journal of Global Information Management*, 10(1), pp5-12, 2002.
- [5] Felix Kaufman, "Data Systems that Cross Company Boundaries, *Harvard Business Review*", *Harvard Business Review*, Jan.-Feb., pp149-161, 1966.
- [6] Barrett, S. and Konsynski B., "Inter-Organization Information Sharing Systems", *MIS Quarterly*, Special Issue, pp93-105, 1982.
- [7] Cash, J.I. and Konsynski, "IS Redraws Competitive Boundaries", *Harvard Business Review*, March- April, pp132-142, 1985.
- [8] Malone T.W., J. Yates and R.I. Bejani, "Electronic Markets and Electronic Hierarchies : Effects of Information Technology on Market Structure and Corporate Strategies", *Communication of the ACM*, Vol. 30, No. 6, June, pp484-497, 1987.
- [9] J.Yannis Bakos, "Information Links and Electronic Marketplaces: The Role of Interorganizational System in Vertical Markets", *Journal of Management Information Systems*, Vol. 8, No. 2, pp31-52, 1991.
- [10] Konsynski, B.R., "Strategic Control in the Extended Enterprise ", *IBM System Journal*, Vol. 32, No. 1, pp112-130, 1993.
- [11] Dietrich, M., "Transaction Cost Economics and Beyond : Towards a New Economics of the Firm", *London: Routledge*, 1994.
- [12] Mahoney, J. T. and Rajendran, P. J., "The Resource-Based View Within the Conversation of Strategic Management ", *Strategic Management Journal*, Vol. 13, pp363-380, 1992.
- [13] Robertson, T.S. & Gatignon, H., "Technology development mode : A transaction cost conceptualization ", *Strategic Management Journal*, Vol.19, pp515-531, 1998.
- [14] Barney, J., "The Debate between Traditional Management Theory and Organization Economics : Substantial Differences or Intergroup Conflict ? ", *Academy of Management Journal*, Vol. 7, pp347-359, 1990.
- [15] Williamson, O. E., "The logic of economic organization", *Journal of Law Economics and Organization*, Vol.4, No.1, pp 65-93, 1988.
- [16] Anderson, J. C. and J. A. Narus, "A model of distributor firm and manufacturer firm working relationships", *Journal of Marketing*, Vol.54, January, pp42-58, 1999.
- [17] Zaheer, A. and Venkatraman N., "Determinants of Electronic Integration in the Insurance Industry: An Empirical Test", *Management Science*, Vol. 40, No. 5, pp549-566, 1994.
- [18] Emerson, R. M., "Power-Dependence Relations ", *American Sociological Review*, Vol. 27, pp31-41, 1962.
- [19] Pfeffer, J. and G. R. Salancik, "The External Control of Organizations", *New York : Harper and Row*, 1978.
- [20] Sengupta Sanjit. Robert E. Krapfel and Michael, A. Pusateri,, "Switching Costs in Key Account Relationships", *Journal of Personal Selling & Sales Management*, Vol.17, Fall, pp9-16.
- [21] DiMaggio, P. & Powell, W.W., "The iron cage revisited: institutional isomorphism and collective rationality in organizational fields" , *American Sociological Review*, pp147-160, 1983.
- [22] Fogarty, T. J., "Organizational socialization in accounting firms: a theoretical framework and agenda for future research", *Accounting, Organizations and Society*, Vol. 17, No. 2, pp 129-149, 1992.
- [23] Galakiewicz, Josph., "Professional Networks and the Institutionslization of the Single Mind Set " *American Sociological Review*, pp639-658, 1985.
- [24] Lachman, R. and Aranya, N., "Job attitudes and turnover intensions among professionals in different work settings", *Organization Studies*, pp279-294, 1986.
- [25] Avgerou, C. "The Significance of Context in Information Systems and Organizational Change ", *Information Systems Journal*, 11, pp43-63, 2001.
- [26] Kumar, K., H. G. van Dissel, and P. Bielli., "The Merchant of Prato -- Revisited: Toward a Third Rationality of Information Systems", *MIS Quarterly*, June, pp199-226, 1998.
- [27] Sharma, S., "Applied Multivariate Techniques", *Wiley*, New York, 1996.

Digital Divide: Adoption of Internet in China

Kanliang Wang, Qing Liu

The School of Management, Xi'an Jiaotong University, Xi'an 710049, China
klwang@mail.xjtu.edu.cn

ABSTRACT

With the development of information and communications technology (ICTs), Internet has been one of the major driving forces of the new economy. Nevertheless, access to the new technology remains extremely unequally distributed across and within societies, and there have a good deal of popular discussions about this "digital divide" in academia and business communities of every society. This study provides empirical evidence on China's Internet development and diffusion in recent years. Findings show that there is a clear digital divide among the nation's three economic zones, and while the growth rate for new Internet users is accelerating and the infrastructure of network is improving, the digital divide is growing narrower. In addition, by undertaking a set of regression analysis, the paper also finds that income and education play a critical role in shaping the divide. Based on the result, the paper offers some related policy suggestions to promote a wider diffusion of internet in poorer regions.

Keywords: Digital divide, Internet, Gini coefficient

1. INTRODUCTION

Information technology is changing the way people communicate, learn and work, and the competitive advantage enjoyed by those having Internet access is becoming increasingly obvious. However, not everyone has access to this technology. The idea of the "digital divide" that put forward by NTIA firstly refers to the growing gap between those who have access to computers or Internet and those who have not access, ranging from the global level, to nation states, to communities, and to individuals [1]. Though these applications have some common features, there are differences on the response of policy and the composition of stakeholders.

Digital divide has been a continuing buzzword in public discourse and international agendas after it had been put forward. However, existing studies have largely been confined to the boundaries of national states. The explosive growth of information and communications technologies (ICTs) in recent years, particularly the rise of Internet and its related applications, has created unprecedented opportunities, but also created threats for developing countries. From the opportunity aspect, some believe that ICTs will be a lever that may advance developing countries to enter into "leapfrogging" seedtime. This perspective is in line with Barlow's (1998) comment that Africa should skip industrialism entirely and leap directly into the information era [2]. This conclusion is shared by Hudson (2001) who suggests that the potential for "leapfrogging" lies in the use of wireless communication technology [3]. Nevertheless, According to Braga (1998), some believe that ICTs have and will contribute to even wider economic divergence between developing and developed countries [4]. This view is consistent with Brown's (2001) argument that there is still a lot of skepticism with regard to whether ICTs can reduce poverty in the developing countries [5]. This idea is also evident in Chowdhury (2000) that some skeptics

still do not see any positive role of ICTs in poverty alleviation. In his words, "the poor can't eat high-speed Internet access, of course" [6]. Opportunity may be outweighed by considerable threat arising from their second-mover position on the adoption of new technologies. Therefore, it is more difficult for an entire nation to leapfrog other nations technologically because of the infrastructure and human capital bottleneck [7]. What most important is that the developed countries have an advantage of exploring new technologies. While the developing countries learn, assimilate and understand original technologies, the developed countries have already applied the advanced ones. Therefore, there is a widening digital divide between the developed countries and the developing countries.

Research on Internet diffusion is accompanied by the evolution of Internet itself. With the increasing use of Internet in developed countries, research about its applications has been on the rise. Most studies have taken place in OECD countries, especially the US, Japan and Western European countries. The Department of Commerce's National Telecommunication and Information Administration's reports, generally entitled *Falling Through the Net*, have been the primary and most consistent way to track computer ownership and network access across broad demographic categories over a relatively long period of time (since 1995) [1]. Even though there is a deeper and wider digital divide, little research has paid attention to how Internet access and use fit into everyday life in developing countries, such as India and countries in Africa [8, 9]. These studies suggest that digital divide has been one of the leading economic and civil rights issues, and call for government action to ensure that all people are able to enjoy the competitive advantages gained by Internet participation. Widespread diffusion does not imply ubiquity, even within developed countries. Digital divide appeared at the very start of Internet [10]. There is the same

challenge in China too. Adoption and development of ICTs drives the development of new economy in China. Meanwhile, there has always been great regional variation in three zones: the eastern, the central and the western zones [11]. From 1990 to 2002, the percent of GDP of the eastern zones contributed to the nation increased from 51.5% to 57.86%. But this percentage in the central and the western zones decreased 3.14% and 3.23% respectively. GDP of the eastern zones was 2.0 and 2.6 times than that of the central and the western zones in 2002 respectively. The Gini coefficient in China, the important index measuring the income inequality, reached 0.41 in 2000, which was well beyond the normal level. These indicate the increase of regional gap in economic development. Moreover, the number of Internet users in the eastern zones was 1.95 and 3.03 times than that of the central and the western zones respectively in 2003. There are some studies on digital divide in China at present. But most of them focus on the qualitative analysis [12, 13, 14]. In addition, there are also few studies on how Internet diffused in China [15]. Empirical researches on digital divide in China are scarce.

By undertaking a set of regression analysis on Internet-related data in China, this paper aims to analyze the status quo of Internet development and diffusion in China, and provides empirical evidence on the digital divide of Internet adoption across China. Moreover, it discloses the relation between the development of economy, education and Internet. These results may offer some implications to the policies that may bridge the digital divide.

2. RESEARCH QUESTIONS

At present, digital divide not only exists in the development of information technology, but also in the application of information technology. So, digital divide has multiple dimensions. This study focuses on the digital divides between users and non-users of Internet in China.

Early characterizations of the divide singled out income and educational attainment as correlating strongly to computer and network access [16]. Thus, although there are strong relations between income and education in this early data (one-third of the income gap can be attributed to education), education (and other variables) cannot completely account for the fact that the wealthy own more computers and have greater access to computer networks than the poor. Income and education considered as separate variables are significant contributors to the access. Among OECD countries, the digital divide between high income households and low income households varies from country to country, ranging from a gap of more than 60% in the UK to less than 20% in Denmark [10]. With respect to the NTIA's historical concerns, the effects of income and education still exist as well: the better educated and compensated

far outpace those of alternate class positions. Thus income and education continue to be the variables that correlate most strongly with Internet access and computer ownership [17]. Whether there is similar relation between these two factors and Internet digital divide of China is an important factor in formulating related policy to solve the problem.

China accesses firstly Internet by its CSTNET at April 1994, which has been a great development in recent years. By reports on Internet development in China (CNNIC, 2004), the number of Internet users is more than 79.5 million, and the number of computers that can access Internet is more than 30.89 million. However, there is serious regional inequality on development and diffuse of Internet. Recent report shows that the number of Web sites in three regions (eastern China, northern China and Southern China) is 90% of all sites across China. The digital divide is widening the gap of among these regions, which could bring serious social problems. Some domestic studies find that there is a regional digital divide. There are great differences in Internet host and Internet users among these regions. In general, the number of Internet users in the rich regions is far excess that of the poor regions [13]. But to what extent is the gap in Internet?

Computer and Internet use at home and work are increasing rapidly, but there is disagreement on whether access is increasing among all groups at similar rates. One analysis suggested that the divide may shrink in the near future because the number of computer users with lower incomes is increasing more rapidly than the total number of computer users overall [18]. In addition, Howard (2001) noticed the dynamics of digital divides, and he agreed that infrastructural access gaps are narrowing, though narrowing in unequal ways [19]. However, other researches have found that the digital divide also continues to grow at an alarming rate. On the one hand, the connected ones - such as those with higher income and better education - are adopting newer technologies faster and are connecting even more. On the other hand, groups with limited Internet and computer access continue to lag far behind [20]. One study concluded that these digital divides are increasing even after taking into account that "adoption of new technologies tends to be fairly slow at low penetration levels, faster thereafter and slower again as it reaches saturation" [18]. Then what direction does the digital divide of Internet adoption in China changed in several years ahead?

Based on above discussion, this paper aims to investigate the following questions:

- Is there any relation between the number of Internet users and the levels of income and education across different regions in China?
- To what extent are there digital divides in Internet adoption among three zones: the eastern, the central and the western zones?

◦ What direction will the digital divide in Internet adoption change across different regions?

3. METHOD

3.1 Sample

The data summarized here emerged from all survey reports on Internet development in China published by CNNIC [21]. In addition, the study refers to relevant data in China Statistical Yearbook and Regional Statistical Yearly Report (2004) [22, 23]. The object of the study includes 31 provinces or regions of mainland China. Thus analysis is done on two aspects: one is the relation between income, education and digital divide across different regions of China, another is the status quo and the trend of the digital divide origins from inequitable diffuse of Internet.

3.2 Variables

As the dependent variable, the digital divide on adoption of Internet is measured by the number of Internet users. According to CNNIC, Internet users include Chinese who accessed Internet for at least one hour per week in average. Income, as an independent variable, is measured by a common index, which is GDP per capita [24]. As another independent variable, education was measured by the number of primary, secondary and tertiary students [24], but these three types of students have a strong relation. It was also measured by the number of people who own senior secondary education [15], but it doesn't take into account of the fluidity of them. To avoid these problems, this paper measures education by the number of student enrolled at colleges and universities. To some extent, the index reflects the level of education in a region.

4. MODEL AND FINDINGS

4.1 The relationship between income, education and Internet users

What relation is there between the number of Internet users and income, education across different regions in China? Simple regression results, as reported in Table 1, suggest the following.

Table 1. The relationship between income, education and Internet users

Variables	Internet users	Income	Education
Internet users	1		
Income	0.931**	1	
Education	0.876**	0.798**	1

Note. *Internet users* is measured by the number of Internet users per 1,000 people, *income* is measured by GDP per capita and *education* is measured by the number of student enrolled at colleges and universities. All data is calculated by a natural logarithm scale.

** Statistical significance at the 1% level.

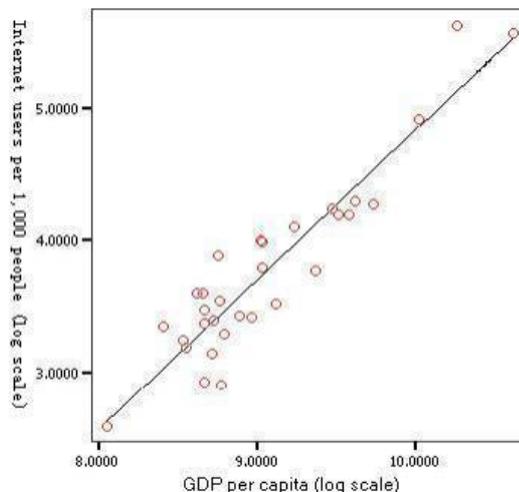


Figure 1. Internet user density and GDP per capita in 23 regions, 2002

Data source: *China Statistical Yearbook and Report on Internet Development in China (2003)*

We observed that the number of Internet users is still highly correlated with income and education. Income is the primary factor influencing the access and adoption of Internet, which possibly origin from the inequality of regional economy development (see Fig. 1). In addition, the level of education has a strong relation with the number of Internet users (87.6%), which can be attributed to two reasons. Firstly, education contributes directly towards basic literacy, and reading and writing skills are essential in using this technology. Well-educated people are also likely to be faster to adopt new innovations than people with little education. Secondly, given that the early users of Internet were people working in higher education and research, academic institutions may play an important role in spreading Internet. They are often among the first institutions in a nation to be wired. Moreover, there is a tight relation between income and education (coefficient is 0.798). The result has a little difference with prior conclusion [15], for the reason of the difference of index to measure education.

4.2 Extent of digital divide

In order to measure extent of the digital divide in Internet adoption among three zones: the eastern, the central and the western zones, the number of Internet users of every region in China for the years 1997-2003 were analyzed. Table 2 summarizes the number of Internet users per 1,000 people for this period in different regions. It shows that there is a clear digital divide in Internet adoption among these three zones, though the adoption of Internet in the central and the western zones has been developing quickly from years 1997-2003. In 2003, the number of Internet users per 1,000 people in the central and the western zones was 1/3 of that in the eastern zones.

Moreover, the increase rate of Internet adoption in the western zones has been higher than that in eastern zones since year 2000, which is highest among three zones. The result maybe attributed to two reasons. One is the support of the central government policy. The central government released the Western China Development policy, which had a great promotion to information infrastructure development. Secondly, there was a great development in western economy. It brought more people to access the new technology.

Table 2. Internet users per capita 1,000 people by zones, 1997-2003

	Eastern	Central	Western
1997	3.05	0.37	0.18
1998	6.10	0.71	0.66
1999	27.71	3.26	2.53
2000	48.44	11.68	11.31
2001	70.25	17.11	15.26
2002	100.19	34.45	31.13
2003	123.65	42.63	45.53

Data source: *China Statistical Yearbook* (1998-2003), *China Regional Statistical Yearly Report* (2004) and *Report on Internet Development in China* (1997-2004)

Fig.2 characterizes the changing status of Internet users per capita in the three zones. It shows that the adoption of Internet is nearly similar between the central and the western zones, but there is a great gap by comparing with that of the eastern zones, and the gap is widening along with time. Therefore, there is a trend that the digital divide on the adoption of Internet is widening.

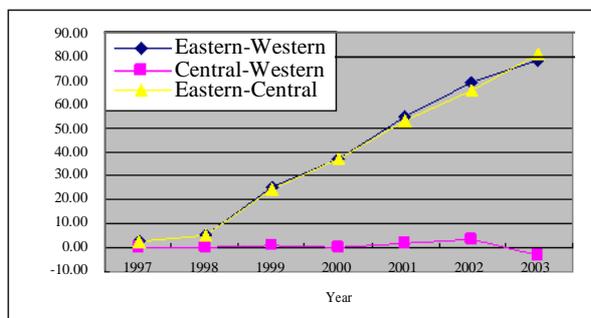


Figure 2. The difference of Internet users between three zones, 1997-2003

4.3 Trend of digital divide

Because there is a gap in regional economy and education, the adoption of Internet has been developing unevenly. With government's regulating the policies of IT, whether the digital divide of regional unequal development is narrowing? By comparing two or more figures, or columns of figures, it can see which one is greater or lower. For example figures for the eastern zones can be compared with that of the central and the western zones to see which one is greater or lower. This is the easiest way to look at disparities. It is a simple way that every one interested can work out and get a feeling

of the level of disparity between different zones. However, it is crude, and may not tell us much. At times it is not possible to scan through all the figures at a time. This section utilizes Gini coefficient to measure the digital divide across regions. Gini coefficient can quantify the level of inequality.

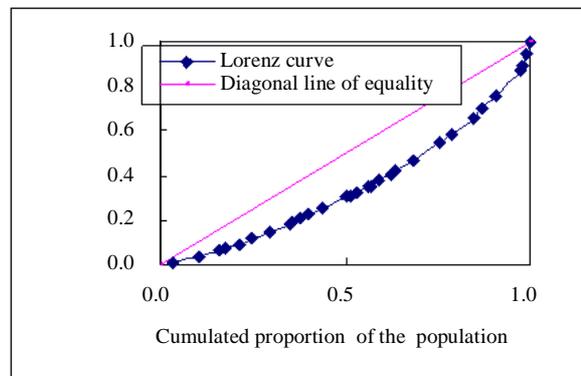


Figure 3. The calculation of Gini coefficient, 2003

The Gini coefficient is based on the Lorenz curve, a cumulative frequency curve that compares the distribution of a specific variable with the uniform distribution that represents equality (See Fig. 3). This equality distribution is represented by a diagonal line, and the greater the deviation of the Lorenz curve from this line, the greater the inequality [25]. In general, scientists still use this coefficient to measure the level of wealth distribution between nations, income between households, health among community etc. When applying this index to Internet variables, the cumulative proportion of the population is generally shown on the X axis, and the cumulative proportion of Internet users on the Y axis. The greater the distance from the diagonal line, the greater the inequality.

The Gini Coefficient ranges from 0 to 1, 0 representing perfect equality and 1 total inequality. It corresponds to twice the area between the Lorenz curve and the diagonal. There are different methods to calculate the Gini Coefficient, but a simple formula, shown below, was provided by Brown (1994) [26].

$$Gini = 1 - \sum_{i=0}^{k-1} (Y_{i+1} + Y_i)(X_{i+1} + X_i)$$

where, Y=cumulated proportion of Internet variable; X=cumulated proportion of the population variable.

Table 3 summarizes the Gini coefficients of Internet for years 1997-2003. Generally, Gini coefficient of Internet has been decreasing. Internet Gini coefficient is 0.422 in 1997, which is twice of that in 2003. It indicates that there was a serious inequality in early period of Internet development. There was a wide decrease form 1999 to 2000. It should benefit from the government policies and intervention of venture capital. Since then, Internet has served Chinese people and enterprises really.

Table 3 Internet Gini coefficients in China, 1997-2003

Year	1997	1998	1999	2000
Gini coefficient	0.422	0.438	0.412	0.289
Year	2001	2002	2003	
Gini coefficient	0.293	0.248	0.215	

Data source: *China Statistical Yearbook* (1998-2003), *China Regional Statistical Yearly Report* (2004) and *Report on Internet Development in China* (1997-2004)

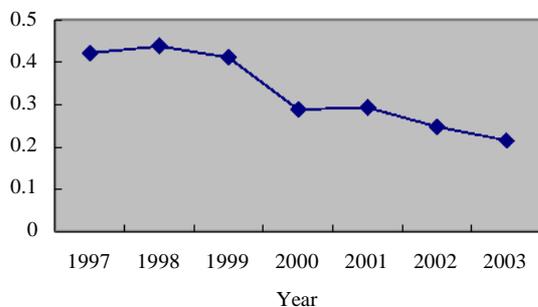


Figure 4. The trend of Gini coefficients in China, 1997-2003

Fig. 4 is based on data of Table 3. It shows that the digital divide on adoption of Internet is narrowing across regions for years 1997-2003.

5. BRIDGE THE DIGITAL DIVIDE

For any nation, the adoption of information technology doesn't stop its steps because of the lagger. The digital divide can be seen as both a symptom and a cause of underlying disparities in society. So whether the digital divide can be bridged quickly and effectively relate to sustainable development of a society, and this requests nation and local government to make great efforts. In this section, some policy suggestions for bridging the digital divide are proposed

5.1 Expediting establishment of info infra-structure

For a region to succeed in Internet area, information infrastructure plays a critical role. That is to say, establishment of information infrastructure is a basis for bridging the digital divide. Government plays an important role in creating such infrastructure, especially in the poorer regions. This role stems from a number of important considerations. First, in very poor regions, the market forces may be weak both because of lack of effective demand and because of lumpiness of investment. Second, in order to attract the private sector to invest in information infrastructure, government may have to play the role of a catalyst by instituting various innovative incentive mechanisms such as build-own-operate, build-own-transfer, etc. Finally, government opens its market to foreign investors, which can speed up transfer of technology and encourages investments on Internet infrastructure. Bridging the digital divide between the eastern and the western zones requires that the establishment of information infrastructure such as computer network and

communication links be expedited. Government should take network as a public establishment. They should offer more chance to access Internet and knowledge about IT for the public. In addition, the establishment of information infrastructure should be an important component of Western China Development Strategy. This can promote the popularization and development of Internet in the western zones.

5.2 Increasing investment in education

As above analysis suggests, the relationship between education and Internet is critical. Education is important because it provides basic skills required for creating, adapting, and utilizing such technologies. This is not to deny that even illiterate or near-literate can possibly take advantage of certain technological applications. But to go beyond elementary applications, education becomes increasingly important. Indeed, international evidence suggests that education is a strong complement, especially secondary and higher education. Therefore, if a region aspires to exploit in significant ways the opportunities offered by Internet, the creation of new industries in particular, it needs to emphasize secondary and higher education.

By the Report on Internet Development in China (CNNIC, 2004), those who have education level below senior secondary education is 13.5% of all Internet users [21]. It indicates that higher education basically provides trainings of Internet skills to students in China, but not for people below senior secondary education. There are two possible ways to deal with it. First, while developing higher education, government should pay much attention to Internet education for primary and secondary schools; Second, in addition to formal education, a rapidly changing technology like Internet would require continuous training on the part of the workforce. However, the government can also play an important role in inducing firms to impart such training through various types of tax incentives.

5.3 Reducing cost

Many countries in the world concern the problem of digital divide, and make every people take the advantage of IT by improving IT, infrastructures and service actively. In China, expense of accessing Internet is too high, which is a bottleneck hampering the popularization of Internet. Though the charge is lower than some developed countries in the world, but considering the level of income, the rate of network expenses to income is much higher, which restrict the diffusion of Internet. So government should develop leasing and low-cost purchasing schemes, and coordinate, develop and identify a range of low-cost technical solutions to Internet access provision, particularly in poorer regions. This would help to reduce the number of people who perceive Internet as unaffordable and increase the effectiveness of using limited resources. Moreover,

government should reduce cost of accessing network by continuing to develop a national infrastructure of low-cost or free access points in under-served regions. Income is one of most important factors affecting the digital divide. The reduced cost of Internet access may attribute to income increase. So reducing cost is one of essential measures to bridge the digital divide.

5.4 Fostering favorable environment

Although the rapid development of Internet requires an open and competitive market, China still maintain strong barriers against entry in this market. The entry barrier, along with the heavy government involvement in such ICT-related sectors as telephones, has spawned pervasive corruption in many regions and has raised the cost of communication. It results in difficult access to Internet. Therefore, to bridge the digital divide and promote the adoption of Internet, government must foster a favorable environment. Open and competitive market is essential, which can ensure enterprises to take part in the establishment of information infrastructure, including private enterprises and foreign enterprises. This can decrease the pressure of government. In addition, governments in poorer countries should foster an environment that would encourage innovation through research and development. It may be noted that developed countries that have established their lead in Internet have invested heavily in R&D activities. With fiscal incentives and proper enforcement of property rights, the private sector would have the impetus to engage in R&D activities. However, in areas where the market fails, particularly in basic research, government should take the lead. Aside from direct investments in R&D, government can encourage linkages between universities and Internet industry in technology transfer.

5.5 Cooperation between government, enterprise and community

In addition to regional efforts, national and international organizations can play a role in promoting Internet in poorer region. It is essential that regional government cooperate with enterprises and society in the establishment of information infrastructure. First, sector of telecommunication in China is generally regulated by government. It is difficult for international organizations to enter into the industry without admission. So central and regional governments should issue deregulation polices gradually, which ensure enterprises free entry of the private sector, including foreign firms. These firms can support a specific set of policy initiatives to bridge the digital divide through technical assistance and policy advice, such as free ISP. Second, regional governments should attract enterprises, multinational companies and venture capital, which is likely to be critical in the growth of the local Internet access. Governments can offer the necessary financial assistance to them.

Moreover, if governments in poor regions channel their

scarce financial and political resources to developing social and human capital, building the basic infrastructure and creating a proper playing field for the private sector, Internet access will flourish and digital divide will diminish. Thus, notwithstanding the concerns voiced on the perils of being left behind in this digital age, poor regions should carefully balance the resource conflict of adopting modern technology and economic development.

6. CONCLUSION

The ability to access internet has become increasingly important to completely immerse oneself in the economic, political, and social aspects of the world. The digital divide is arguably an indicator of deeper divides within society. It can be seen as both a symptom and a cause of underlying disparities in society. This paper investigates the adoption of Internet across different regions in China. The Gini coefficient is calculated using data on Internet users per capita for the years 1997-2003. The basic finding is that there is a clear digital divide about Internet adoption across regions, and the divide has grown narrower. In contrast with some earlier findings (Bikson, 1999; Besser, 2001), these access gaps are narrowing [18, 19]. This paper also finds that the divide is still highly correlated to income and education, the conclusion is consistent with early result (Grabill, 2003) [17]. Based on this analysis, the paper also offers some policy suggestions as to how to promote a wider diffusion of Internet in poorer societies.

There are at least two important avenues for future researches. One avenue is to analyze the factors affecting the adoption of Internet except for income and education, such as cost and race etc. This will help to understand the digital divide deeply. Another avenue is to study in greater detail the adoption of the cluster of Internet-related technologies, because digital divides on other technological aspects have an important effect on society.

ACKNOWLEDGEMENT

The research was partly supported by the National Natural Science Foundation of China (grant number 70121001 and 70372049) and 211 project.

REFERENCES

- [1] *Falling Through The Net: Defining The Digital Divide*. National Telecommunications and Information Administration, U.S. Department of Commerce, 1999.
- [2] Barlow, John Perry (1998). Africa Rising: Everything you know about Africa is wrong. *Wired*. Jan. 1998.
- [3] Hudson, H. E. (2001). The Potential of ICTs for Development: Opportunities and Obstacles. *Telecommunications Management and Policy*

- Program*. University of San Francisco. *on systems Engineering*, Vol. 1, No. 1, pp35-42, 1994.
- [4] Braga, C. (1998). Inclusion or exclusion? *UNESCO Courier*. Retrieved May, 2004. http://www.unesco.org/courier/1998_12/uk/dossier/tx t21.htm.
- [5] Brown, M.M. (2001). Can ICTs Address the Needs of the Poor? A Commentary from UNDP. <http://www.undp.org/dpa/choices/2001/june/>. Retrieved May, 2004.
- [6] Chowdhury, N. (2000). Information and Communications Technologies and IFPRI's Mandate: A Conceptual Framework. Retrieved May, 2004. <http://www.ifpri.org/divs/cd/dp/ictdp01.pdf>.
- [7] Jalava, J., Pohjola, M., (2002). Economic growth in the new economy: evidence from advanced economies. *Information Economics and Policy*, 14 (2), 189-210.
- [8] Singh, Jagtar (2002). From Atoms to Bits: Consequences of the Emerging Digital Divide in India. *International Information & Library Review*. 34(2), 187-200
- [9] Roycroft, Trevor R.; Anantho, Siriwan (2003). Internet subscription in Africa: policy for a dual digital divide. *Telecommunications Policy*. 27(1-2), 61-74.
- [10] Chen, Wenhong; Wellman, Barry (2003). Charting and Bridging Digital Divides. *I-Ways*. 26(4), 155.
- [11] Meng, Qingxuan; Li, Mingzhi (2002). New Economy and ICT development in China. *Information Economics and Policy*. 14(2), 275-295.
- [12] FANG Jiaping. Digital Divide (In Chinese). *Scientific Decision*. 2001(12), 40-42.
- [13] CHEN Ying. Digital Divide of Internet in China (In Chinese). *China Information Review*. 2001(11), 35-36.
- [14] YANG Lin, LI Mingzhi. The Present Situation and Countermeasures of Digital Gap among Different Regions in China (In Chinese). *China Soft Science Magazine*. 16(4) 2002, 22-26.
- [15] JIN Jianbin. Diffuse of Internet in China Mainland (In Chinese). APSP2001 Symposium, Chinese Communication Society. 2001.
- [16] National Telecommunications Information Administration (1995, July). Falling through the Net: A survey of the "have nots" in rural and urban America. Retrieved May, 2004. <http://www.ntia.doc.gov/ntiahome/fallingthru.html>.
- [17] Grabill, Jeffrey T. (2003). On divides and interfaces: access, class, and computers. *Computers and Composition*. 20(4), 455-472.
- [18] T.K. Bikson and C.W.A. Panis. Citizens, Computers, and Connectivity: A Review of Trends. *Santa Monica, Calif.: RAND*, 1999.
- [19] Besser, Howard. (2001). The next digital divides. Teaching to change LA. Retrieved March 28, 2004. <http://www.tcla.gseis.ucla.edu/divide/politics/besser.html>.
- [20] Latimer, CP (2001). Government Information Focus. The Digital Divide: Understanding and Addressing the Challenge. New York State Forum for Information Resource Management. Retrieved May, 2004, from <http://www.nysfirm.org>.
- [21] CNNIC. Reports on Internet development in China (1997.10-2004.1). <http://www.cnnic.net.cn>. 2004.5.
- [22] National Bureau of Statistics of China. China Statistical Yearbook (1998-2003). <http://www.stats.gov.cn>. 2004.5.
- [23] National Bureau of Statistics of China. Yearbook and Regional Statistical Yearly Report (2004). <http://www.stats.gov.cn>. 2004.5.
- [24] Quibria, M.G.; Ahmed, Shamsun N.; Tschang, Ted; Reyes-Macasaquit, Mari-Len (2003). Digital divide: determinants and policies with special reference to Asia. *Journal of Asian Economics*. 13(6), 811-825.
- [25] C. Castillo-Salgado, C.Schneider and E. Loyola, etc. (2001). Measuring Health Inequalities: Gini Coefficient and Concentration Index. *Epidemiological Bulletin*. (22)1, 3-4.
- [26] M. Brown (1994). Using Gini-Style Indices to Evaluate the Spatial Patterns of Health Practitioners: Theoretical Considerations and an Application Based on Alberta Data. *Social Science & Medicine*. 38(9), 1243-12.

E-business and Organizational Learning: A Case Study of Three Globally Operating Companies

Marjatta Maula

Institute of Business Information Management, Tampere University of Technology, Tampere, Finland
marjatta.maula@tut.fi

ABSTRACT

This paper analyzes how the web-based e-business solutions of three globally operating companies have served their organizational learning. The case companies represent metal and mining, pulp and paper machinery and pharmaceuticals industries. It can be concluded that while the e-business solutions may have been part of larger strategic development process, they were not originally designed to serve organizational learning. During the utilization phase, the learning opportunities had remained quite low or underutilized. However, all case companies had learned considerably as a by-product especially during the design and implementation phase. The study indicates that e-business solutions provide plenty of learning potential that should be taken into account in the design, implementation and utilization phases.

Keywords: e-Business, organizational learning, knowledge management, knowledge creation.

1. INTRODUCTION

Web-based e-business solutions enable efficient and intensive interaction with customers. They make possible to create personalized contents and various methods for producing, exchanging and collecting information. They help organization to reach its customers across time zones and geographic boundaries. These technologies also enable profitable solutions for a limited number of big customer firms, or for an unlimited number of individuals all over the world. The major motivation for firms to implement e-business solutions is to increase efficiency and service capability and to enable new kinds of business processes. It could also be assumed that web-based e-business solutions are used to facilitate organizational learning, knowledge creation and renewal that are increasingly important aspects for the firms' competitiveness. Firms could learn from their customers and markets by using information and knowledge that can be channeled through e-business solutions (Skyrme, 2001). However, it is not clear what are firms' real motivations, opportunities and methods to utilize e-business solutions for learning purposes. For this reason the research project "Strategic Management of Innovativeness, Renewal Capability and Knowledge, and the Role of e-Business" was established to study through selected case studies the following research question: "How does a company learn and create new knowledge by utilizing e-business solutions?" The purpose of the study was to create new knowledge and to facilitate learning from their experiences.

2. THEORETICAL APPROACH

The theoretical framework consists of an interpretation of the theory of self-producing systems (autopoiesis theory; Maturana and Varela, 1980;

Mingers, 1995). The interpretation, living composition[®], emphasizes the role of 'boundary elements'. They consist of various roles, functions and systems that enable the firm's interaction with its business environment, help to absorb and create knowledge, and facilitate organizational change, learning and co-evolution with the business environment (Maula, 2000). The web-based e-business solutions are studied here as such organizational boundary elements.

3. EMPIRICAL APPROACH: WEB-BASED BOUNDARY ELEMENTS IN THREE CASE COMPANIES

The empirical part of the study was conducted in 2001-2003 in three globally operating case companies that were selected among internationally recognized frontrunners in the strategic application of e-business:

- Outokumpu Copper Products OCP Europe – EPC Electrical Power and Components Unit.
- Metso Paper
- A pharmaceuticals company.

The three companies differ considerably from each other in their e-business solutions.

3.1 Outokumpu Copper Products OCP Europe – EPC Electrical Power and Components Unit

The core businesses of *Outokumpu Group* are stainless steel, copper products and technology. The company has customers in a wide range of industries: catering and households, building and construction, transportation, industrial machinery and equipment and electronics and communication. The company has 19,000 people in more than 40 countries. The Group's net sales is EUR 6 billion. *Outokumpu Copper* is one of the core businesses. Its net sales

cover 26% of the Group's sales. Outokumpu Copper is the second largest producer in the world in fabricated copper products. The company employs 7,600 people. It has production plants in 15 countries. The products include anodes, brass rod, profiles, rods & bars, sheets & plates, strips, tubes, welding products and wire. *EPC Electrical Power and Components Business Line* employs 400 people. It has 6 production plants in Finland, England and Austria.

EPC Electrical Power and Components Business Line has created a business-to-business extranet solution for its selected customers. Customer relationships are based on long-term contracts, and interaction with them is complex in nature. EPC Business Line started in 1995 to implement the New Sales Mode solution that included a salesman tool box and back office tools. Based on the experiences and new technological and business opportunities the company started in 1998 the development of the New Sales Mode Web (NSMWeb). The objective was to develop a Key Customer Relationship Management and Service System. The development of NSMWeb was a rapid process. It was part of a profitability development program and strategic masterplan that included both strategy specification and operational improvement. It was based on an analysis of the customer contacts and the work of sales persons.

One of the points of departure for the NSMWeb development was the general development of marketing in Outokumpu Group. It meant a change from production oriented to sales oriented and further on to customer oriented marketing methods. Moreover, at EPC the strategy was specified (vision remained the same) and operative systems, such as order/delivery process, were developed simultaneously. Customer management had high priority, and the sales processes were to be reorganized. The e-business solution was built on existing technology (Lotus Notes) and databases. It was to provide "bullet proof reliability" and improve the image of the firm.

The solution enabled the design of customized business processes. The objectives of the solution was the inquiry & order management for selected key customers. The purpose was to create tailored services and to generalize them later. The development occurred in intensive co-operation between the producer and customer and it aimed at win-win solutions and learning.

One of the major objectives included the transformation of routine sales work into technical services. As a result the roles of the salespersons were radically changed into expert role, and they were supported by NSMWeb solution and incentive bonus plans. The sales work, customer contacts and

acquisition of new customers covered earlier only 20 % of the time of the sales persons. The objective was to increase it up to 60-70 % of their time, and the solution reached quite successfully these objectives.

The development of the NSMWeb solution helped to understand the real needs of the customers especially during the design and implementation phase. EPC could intensify the discussion with the customers, which evoked ideas of new services and products, and new methods to speed up the deliveries. It improved the communication with the customers and made it deeper.

The new roles of the salespersons and the web-based solution constitute thus two kinds of boundary elements that have strengthened customer relationships and enabled efficient order management and integration to scheduling and manufacturing. They also improve customer service, knowledge management, and learning and renewal capability of the firm. The new role definition and NSMWeb solution are examples of boundary elements that help to manage internal and external complexity better. NSMWeb was later developed into the One Common Application.

3.2 Metso Paper

Metso Paper is a business area of Metso Corporation that is a global supplier of process-industry machinery and systems, know-how and aftermarket services. Metso has business operations in more than fifty countries. In 2002, it had 17 technology centers, 30 manufacturing units, 32 service centers and 25 sales subsidiaries. With corporate net sales in 2002 of 4.7 billion euros, Metso had approximately 30,000 employees.

Metso Corporation has as its strategic objective the transformation of the corporation into a networked and knowledge-based technology company. *Life cycle technology*, enabled by the Metso Future Care business concept, forms a central part of this strategy. Life-cycle technology is an umbrella term for RTD (research and technological development) projects based on process life cycle knowledge. It aims to improve the customers' profitability by managing the product life cycles and product-related knowledge. *Metso Future Care* provides a powerful framework for corporate-wide knowledge management and organizational learning. It is a business concept for creating new market potential. It aims at creating new ways of managing customer relationships and developing earnings logic by exploiting previously installed machines and delivered processes to add to the competitiveness of the customer. Over the years, Metso has marketed to its customers 2,000 paper machines, 800 pulp systems, 30,000 crushers, 15,000 screens and 3,000 grinders. In 2002, the main focus

was on after-sales services, remote diagnostics in maintenance, support of the customers in preparing investments, and customer relationship management.

Metso Paper (fiber and paper technology) is the world's largest paper and tissue making line supplier. It covered 38 % of Metso's net sales in 2002. It also has a strong position as a supplier of pulping equipment and systems for the fiberboard and converting industries. In other words, Metso Paper designs and manufactures pulp and paper industry processes, machinery and equipment. Moreover, it provides expert services for developing customers' production processes. Metso Paper has several engineering and manufacturing locations and customers all over the world. Metso Paper's business lines include paper technology, pulping technology, and other businesses.

Metso's e-business solutions can be seen in a broader framework in which both interaction with customers and individual transactions form part of a corporate-wide strategy, customer-wide life-cycle and supplier-wide partnership. In-depth know-how concerning the customers' industrial core processes across the businesses, and the development of value-enhancing solutions necessitates an efficient reciprocal boundary (interface) between the customers and Metso. Specifically, the portals play a central role in organizational learning and knowledge management. The portals not only support internal and external users but also facilitate interaction with customers and suppliers.

Metso Portal is an extranet service that supports the businesses of pulp and paper industry customers. It is a coordinated Metso-wide project, providing an easy doorway to the services of Metso Paper, Metso Automation and Metso Drives through a single log-in. Metso Portal is not a marketplace, but rather enables interaction with the customers, complements other customer-contact channels, and provides a platform for new services. The portal service was started in 2001, and by the beginning of 2002 had over 400 customers using the service (Suvanto, 2002).

PartnerWeb is a tool for the engineering and manufacturing partner network, initiated by materials management. Technically, it utilizes Notes databases and web browsers. PartnerWeb provides several benefits. It improves the flow of information and allows the follow-up of delivery schedules. Engineering and manufacturing instructions are easier to update. PartnerWeb also improves the processing of revisions and claims. Process costs have been reduced at both ends, and document turn-around time has become shorter.

The new *remote diagnostics* centers in Finland, Sweden, and the USA serve pulp and paper industry customers globally. The remote diagnostics solution

involves monitoring, measuring and analyzing various aspects of the processes and machinery with ICT technology from a distance. Service is no longer bound by location. Remote diagnostics tools provide the following benefits: efficient utilization of expertise, reduced traveling, more efficient working time and screening of the essentials from the data.

Especially the portals have contributed to organizational learning and knowledge management at Metso Paper. The implementation phase has provided opportunities to learn about the customers. Also during the utilization phase they facilitate the accumulation, exchange and creation of knowledge. Information about the customers is collected in databases and utilized interactively by the customers and the corporation. Metso's portals help to develop new products, operations and market position. The PartnerWeb supports internal and external users and facilitates interaction with suppliers.

3.3 The Pharmaceutical Company

The Pharmaceutical Company is a globally operating healthcare company that manufactures and markets pharmaceutical products and services. The company has production facilities in almost ten countries, and affiliates or offices in about 70 countries. The Pharmaceutical Company reaches out to almost 200 countries around the world. The company has about 20,000 employees. It is a world leader in the care of one serious illness and also has a leading position in some other pharmaceutical product areas. The illness can occur at all ages and it is becoming increasingly common in developed and developing countries.

The company has implemented comprehensive web-based solutions that are centrally coordinated, locally maintained, and enable interactive communication between the company, consumers and professionals. Development of these solutions has resulted in much of the tacit knowledge becoming explicit. The e-business strategy of the company consists of three focus areas: (1) the websites for patient communities and professionals, (2) corporate communication, which covers the corporate site and all the affiliates, and (3) the websites concerning the products. The focus in this paper is on the first focus area. The web solutions have facilitated interaction with consumer and professional communities. Many parties outside the company have thus become more firmly involved in the learning processes and have gained the opportunity to improve their knowledge.

Although the content of the solutions is basically local, it is based on shared, centrally created structures and principles. The strong orientation towards localization allows for differences in culture and lifestyle, regulations, logistics, and other relevant aspects.

The websites for patients and patient communities are open to anyone in about 20 countries. They do not include any recommendations for the patients. Instead, they provide patients with tools to keep better track of their condition, such as 'Living with the illness' sites for sharing experiences, personal profiles, diary and scorecard for keeping track of the development of the person's own illness. All recommendations should come only from their own personal physician. The solution includes many kinds of applications that provide information for the patient. Examples of them are a library, dictionary (the medical words people don't understand), 'ask the experts' sites, an expert panel, a game for smaller children, and a jeopardy about the illness.

The "Illness information for the professionals" website is offered as a service from the company to health care professionals, such as physicians, nurses, pharmacists, and trainers. The site contains professional information and research, such as up-to-date medical information, educational resources, and the latest research in the illness. An editor reads through medical journals and creates a short calendar describing events and current debate within the field. An abstract database allows physicians to search in journals for abstracts. The site also includes praxis tools and a monthly newsletter co-edited by an in-house doctor, a medical advisor.

The Pharmaceutical Company is a knowledge-intensive company. The success of the company is based on the development and application of explicit scientific knowledge. However, also market knowledge and the understanding of the customer needs are important for the success of the firm. The development and maintenance of the websites have facilitated organizational learning in several ways on local and global levels. The development of the products occurs through a long and complex procedure overseen by R&D specialists. Through the web solutions, The Pharmaceutical Company has become more aware of its business environment and of many changes occurring within it. During the implementation of the web solutions, a large amount of knowledge that was previously in tacit form has now become explicit. This has facilitated organizational learning at the coordinating and local levels. For example, the creation of community sites has made the eBusiness Department and the affiliates more sensitive towards the different perspectives of the various customer groups and professionals, and of the cultural and regulatory differences.

The Pharmaceutical Company's web solutions also facilitate interaction within the consumer and professionals' communities, and promote their educational aim. These solutions have thus allowed The Pharmaceutical Company to reach many parties outside the company, to more tightly connect them to

the learning process, as well as provide these stakeholders with an opportunity to improve their knowledge. The websites serve as interactive boundary elements that help to distribute knowledge between the firm, user communities and professionals, as well as to monitor and collect information. The solutions facilitate the utilization of experts and opinion leaders within and outside the company.

The sites provide access to an extensive amount of data. However, the capability of the solution has not been fully utilized. Many regulations prevent the firm from using the information and knowledge produced by the system. Even that information which could be used within the confines of the regulations has not been very efficiently utilized. The legally available data has great potential as a source of organizational learning on global and local levels for market analysis, product development and other learning purposes. This potential has remained underutilized mainly due to the absence of a shared explicit framework or concept of organizational learning. Such a framework would help to coordinate various processes, knowledge resources and ICT systems to facilitate the learning process.

4. COMPARISON AND SUMMARY

The results of the research project are summarized in Table 1 in the last page of this paper.

This paper analyzes how three globally operating companies have utilized their web-based e-business solutions for organizational learning. The case companies represent metal and mining, paper machinery and healthcare industries. All companies are among global market leaders in their industry. They differ from each other as to the type of customers (industrial, consumers) and their number (tens, hundreds, hundreds of thousands / millions). The solutions depend, for example, on the type and life-span of the product. In the metal industry company the sales is based on annual agreements with relatively few clients. Each individual sales activity of the products requires complex calculations and expertise. The e-business solutions are targeted to facilitate the sales process and to provide information about the products to the customers.

In the paper machinery company the products are very complex technical systems that may consist of hundreds of thousands of parts. The life-span of the product may cover tens of years. There is increasing need for faster design and delivery of the product and many kinds of services. The purpose of the e-business solution was to enable extended collaboration between the company and its customers and partners. The collaboration was enabled through a network centered around the case company as a business hub. Technically, the hub was based on the Internet and it

connected the customers' and partners' devices and solutions with the company's own systems and networks. Life-cycle thinking was developed on the corporate level to highlight the need for long-term partnerships, continual and accumulating learning, and the transformation of reactive services towards preventive and further to proactive type. The corporate Internet portals supported remote diagnostics, proactive services and integrated business solutions.

In the pharmaceuticals company the products are sold to millions of customers. However, the sales depend on various kinds of medical experts. The patients use the medicine daily, normally throughout their lifetime. The treatment of the illness requires education of the patients and experts. Because of many local legal restrictions that vary country by country the e-business solution cannot be used as a direct sales channel. Instead it is used mainly for education, for providing information and creating communities among patients and professionals. The solution is interactive, but the legal aspects restrict the storing, analysis and utilization of the data for learning purposes.

While the development of e-business solutions may have been part of a larger strategic development process, they were not originally designed to serve organizational learning and knowledge creation. The solutions targeted mainly to providing information to the customers, not to providing information about customers and market to the case companies. However, the study shows that web-based e-business solutions can create many opportunities for organizational learning. All three companies had learned considerably as a by-product. Learning had occurred mainly during the design and implementation phase because then the company was compelled to transform tacit knowledge about the products, company itself and the customers into explicit form and to share it within the company and among various partners. As to the utilization phase of the e-business solution, the learning opportunities had remained quite underutilized. Information that is channeled through the e-business system is used mainly for operational purposes. It is not consciously collected, analyzed and transformed into new organizational knowledge.

The study indicates two major aspects that could be taken into consideration. First, the emerging learning opportunities during the design and implementation phase could be utilized more consciously. In the case companies the learning occurred as a by-product. Second, the e-business solutions could be designed so

that they help to absorb new knowledge about the market and customers continually and to transform it into organizational learning, renewal and competitiveness.

5. BENEFITS OF THE RESEARCH PROJECT

The research project analyzes the role of e-business solutions from new angles, the ones of organizational learning, renewal and knowledge creation. These aspects have not been the main objectives for the firms when planning the solutions. The study also provides empirical evidence about boundary elements, and thereby helps to develop further the theoretical and methodological living composition[®] framework.

REFERENCES

- [1] Maula, Marjatta "The senses and memory of a firm – implications of autopoiesis theory for knowledge management." *Journal of Knowledge Management*, Vol. 4, No 2, pp. 157-161, 2000
- [2] Maturana, Humberto R. and Varela, Francisco J. *Autopoiesis and Cognition. The Realization of Living*. D. Reidel Publishing Company. London, 1980.
- [3] Mingers, John, *Self-Producing Systems. Implications and Applications of Autopoiesis*. Plenum Press. London. 1995.
- [4] Skyrme, David J. *Capitalizing on Knowledge. From e-business to k-business*. Oxford: Butterworth Heinemann, 2001.
- [5] Suvanto, Pertti, "Metso Portal otettu hyvin vastaan". (Metso Portal Has Been Received Well.) *Intermetso* No.1, pp. 16-17, 2002.
- [6] Company related material, annual reports, etc.

INTERVIEWS

Juhani Horelli, Manager, Portal Solutions, Metso Paper

Simo Nivala, Vice President, IT, Metso Drives

Panu Routila, Managing Director, Alteams Group. *Earlier*: Controller, OCP Drawn Products Unit (Chicago) and Managing Director, OCP Electrical Power & Components EPC Business Line

Ritva Toivanen, E-Commerce Manager, OCP Europe- Electrical Power and Components EPC Business Line, Outokumpu Poricopper

Ismo Rautiainen, President, Telecommunications and Electronics, Perlos Corporation. *Earlier*: Managing Director, OCP Drawn Products Unit (1995-1997) and Business Development Manager, Outokumpu Copper (1999-2002)

Web-specialists of the Pharmaceutical company.

Copyright © Marjatta Maula. All rights reserved.

E-Business Evaluation in Australia: A Discussion Of Some Preliminary Findings

Mohini Singh, John Byrne

School of Business Information Technology,
RMIT University, GPO Box 2476V
Melbourne 3001 Vic Australia
mohini.sing@rmit.edu.au and john.byrne@rmit.edu.au

ABSTRACT

This paper reports on research carried out into e-business evaluation in an Australian context. The main research instrument was a questionnaire based on a similar study in the United States, which was adapted to give an Australian dimension. The survey was sent to e-business managers in a random sample of Australian companies. The paper reports some preliminary results from the analysis of the survey data. The results identify the more important e-business drivers, transaction capabilities, impacts on financial performance and improvement; and problems experienced during implementation.

Keywords: E-business evaluation, B2B e-business, B2C e-business, e-business improvements, e-business drivers

1. INTRODUCTION

Australian organizations have invested heavily to leverage the Internet and transform their traditional businesses into e-businesses in the last seven years. E-businesses are generally divided into the two main models of B2C and B2B e-business. In Australia B2B e-business was worth 1.1% of the gross domestic product in 2001, and B2C e-business was worth 0.17% [15]. Business organisations doing business on the Internet with digitised business processes are expected to achieve business improvements from reduced operation costs, labour, time and paper based errors. E-business investments are claiming a substantial share of overall IT budgets in most organisations based on anecdotal evidence that organisations achieve unprecedented benefits by leveraging the Internet as a medium of business. According to Kearney research report [12] e-business budgets in Australia are about 27% of overall IT budgets. However, senior managers are increasingly under pressure to justify e-business costs. Do these investments pay off? And if so, how can a company make sure they keep paying off? Managers in successful companies struggle to articulate where such benefits come from, and those without success strive to find a way to turn things around and cash in on e-business initiatives. Key value drivers for e-business in the USA have been identified by Barua et. al., [4] to be system integration, customer orientation of IT, supplier orientation of IT, informational (quality, supply continuity, and relationship management) and transactional; internal operation of IT, customer related processes, supplier related processes, customer e-business readiness and supplier e-business readiness. However, it is difficult to predict that the same set of factors is applicable to organisations in Australia and other parts of the world.

E-business research data in Australia to date is mostly on the number of business organizations trading electronically, type of e-business applications, potential benefits of e-business and the application of the Internet to certain business processes [15, 12]. These reports indicate a substantial increase in the uptake of e-business and Internet applications by Australian organizations. Earlier research in Australia [19] highlighted the need for e-business metrics to evaluate benefits. In the year 2001 a media publication [23] reported huge losses incurred by large Australian organizations such as Fosters and the National Australia Bank from their e-business projects. This research was therefore initiated to evaluate the financial and operational performance of e-business and to quantify the success of e-business in Australia. It was accomplished via an online survey and data was statistically analysed. This paper includes a brief review of literature on e-business evaluation, research methodology, analysis techniques, a discussion of e-business performance in Australia and presents some preliminary results from the survey.

2. LITERATURE REVIEW

New technologies such as the Internet and the World Wide Web have made a profound impact on all businesses in Australia and around the world. E-business enables organisations to reduce costs, increase demand and create new business models. It has the potential to benefit all consumers through reduced prices and improved products and information flows [10]. Small and large firms alike can access the Internet and exploit near-zero marginal costs of distribution for their products [10]. Although e-business has been proved to be popular with large business enterprises, small and medium companies also create value by marketing and selling goods and services electronically [9]. Each company is constrained by the amount of graphics and design

capability that the Internet can deliver, so everyone starts from the same position with their Web sites. Australian organisations adopted the B2C e-business model to increase market share, offer better customer service and to reach out to customers at greater geographic distances [19]. Developments in B2B e-business in Australia has seen businesses and the government, both at the State and Federal levels adopting Web-based e-procurement to achieve volume purchase, dealing with a wider choice of buyers and suppliers, lower costs, better quality, improved delivery, and reduced paperwork and administrative costs [22]. E-business investments are claiming a substantial share of overall IT budgets in most organisations to capitalise on the benefits of e-business [12]. Benefits of e-business as outlined by [6, 21, 24] are increased revenue from enhanced sales; reduced marketing costs with online advertising, reduced time in customer service and online sales; supply chain cost reductions from reduced inventory levels, increased competition from suppliers and shorter cycle time in ordering; and reduced administrative costs from automated routine business processes, order confirmation, accuracy of data and an improved competitive position. Other non quantifiable improvements achieved from e-business include a better corporate image, improved communication with customers and business partners and customers via electronic templates, a faster product development lifecycle enabling faster response to market needs, improved customer service, better information and knowledge management, ability to incorporate positive feedback from customers to enhance sales and ability to incorporate intelligent applications of software for data mining and for forecasting trends and demands [19, 6, 24]. Amit and Zott [2] advocate that business conducted over the Internet in the 21st century with its dynamic, rapidly growing and highly competitive characteristics promises new avenues for the creation of wealth. E-business models, methods and the volume of digitisation vary from industry to industry and from organisation to organisation depending on their size, nature of business, technology capability and in-house technical expertise. Although the value of adopting e-business has been recognised, actual achievements from it are not known. Most e-business reports [15, 1, 12] provide an understanding of the level of e-business adoption. Returns from e-business implementations have not been formally evaluated. Grey et al [11] are of the opinion that much of the value associated with e-business comes not only from improvements in the technological infrastructure but from business and organisational transformations. They explain that a critical part of creating business value is identifying the processes to transform and selecting the right initiatives to enable the transformation. The IT infrastructure capability including speed, flexibility, capacity, efficiency, resilience, and security determine the type of applications that can be run and their performance. These applications affect the accuracy, speed and productivity of these business processes in various functional areas of the organisation which have an impact on the overall business

performance of the enterprise. E-business payoffs are generally assessed as IT payoffs under the themes of metrics, environment, technology and processes [13]. As suggested by Shi and Daniels [18] success in e-business includes functionality, integration and scalability, and an evaluation of e-business applications is necessary for further improvements, management strategies and the deployment of technological developments. Devaraj and Kohli [18] emphasise that it is necessary to determine the strategic role of IT in the organisation as compared to other projects and to get an overall economic picture. They also emphasise that IT projects have less apparent and longer payoff duration and IT metrics according to Devaraj and Kohli include profitability, productivity and customer value while e-commerce payoff measures address efficiency, effectiveness and innovation strategy measured along five dimensions of time, distance or geography, relationships, interactions, and product or service. IT evaluation is similar to benefits realisation as suggested by Ashburton and Doherty [3] and that IT evaluation should be explicitly concerned with the on-going management and direction of the project, suggesting that benefits have to be managed if they are to be collected. They also emphasise that evaluation should be concerned with assessing the process of systems development as well as its product so that the systems development process can be improved over time, and that evaluation should be performed as an on-going process. Cronholm and Goldkuhl [7] describe the strategies for Information Systems evaluation to be goal-based evaluation, goal-free evaluation and criteria-based evaluation. Goal based evaluation measure the IT system based on explicit goals from the organisational context. Goal free evaluation is an inductive and situational driven strategy, while criteria based evaluation means that some explicit general criteria are used as an evaluation yardstick.

Although all of the above IT evaluation issues discussed similar implications for e-business, the only research that evaluated e-business performance exclusively at the time this project was undertaken was by Barua et al [4], on which research discussed in this paper is based.

3. RESEARCH METHODOLOGY

This research was accomplished via mixed mode ie online survey method and a mail survey. Online surveys, were considered to be the apt method of investigating e-business organizations since it is technology based, quick, convenient, enables unlimited reach, seeks a response to all questions, responses are downloadable into a database and transportable to statistical packages for analysis. However, due to a disappointing response to online surveys, a mail survey was implemented to complete the research project.

A set of questions to evaluate the performance of e-business in Australia was initially developed in MSWord. Some of these were adopted the USA[4]. The

questionnaire was divided into sections and included questions presented on Likert Scales, as 'yes' and 'no' answers and some fill in the blanks. Respondents were expected to provide answers with radio buttons, choosing an option from the drop down menus or filling in a word or phrase in the space provided. The whole questionnaire was presented in six HTML pages. At the end of page one a respondent was asked to select the 'submit' button and proceed to the next page. The questionnaire was designed so that a respondent could not proceed to the next page unless an answer to all questions was provided. On submission of first page, HTML codes were included to generate a tracking number enabling the respondent to complete the rest of the questionnaire at a later time or date. The responses to the questionnaire were transferred to a database created in MySQL. Each page in the questionnaire was represented as a table in the database, and each response to a question was recorded as an element in the table. The database was designed to store both numeric and alphanumeric data. The questionnaire was sent to a random sample of 725 companies obtained from a database Business Who's Who <http://bww.dnb.com.au/default.asp>. The database was sorted using different criteria to ensure that they were e-business organizations and included the top 500. The questionnaire was disseminated via emails addressed to the e-business manager. A short explanation of the objectives of the research and the URL for the survey was included in the package.

Response Rate

At the end of one week 32 valid responses were received. Therefore a hardcopy of the questionnaire was sent by mail to elicit responses from the same organisations. In the month following the mail out of the survey, online responses increased to 91, and valid hard copy responses received were 78. This research analysis is therefore based on a response rate of 23.3 %.

Data Analysis

To complete the analysis, we grouped participants who responded with 5, 6 or 7 on Likert Scales as agreeing to the proposition proposed in the statement and calculated the overall percentage of participants agreeing to each proposition and the percentage agreeing in Small, Medium and Large enterprises. In order to assess whether there were significant differences between the level of response to a proposition from the Small Medium and Large enterprises and the overall responses, we carried out a statistical test. In the test we assumed that the number of participants, X , followed a Binomial (n, p) distribution, where n is the number of respondents and p is the proportion of respondents agreeing to the proposition. In this paper, we use Binomial Confidence interval developed by Byrne and Kabaila [5]. This interval has the proven properties of increasing endpoints, is exact and cannot be shortened without violating the minimum coverage requirement.

4. FINDINGS

Findings of this research are presented in this paper reflecting the development, application and achievements of e-businesses in Australia. Responses to part 1 of the questionnaire reflect e-business adoption and development in Australia. From this research it is apparent that in Australia e-business is adopted by organisations of all sizes small (32%), medium (34%) and large (34%). Project leaders responsible for the implementation and development of e-business are largely IT managers (22%). Other project leader positions were e-business managers (14%), managing directors (11%), marketing managers (8%) and other middle managers (45%). It is important to note that most e-business project leaders in these organisations generally held postgraduate (30.8%) and graduate (37.2%) level qualifications. All Australian industries have some form of e-business in their organisations although it is most widely adopted by the manufacturing (24%) and the service industry (22%) sectors. Others were grouped as transport/utility (18%), retail/wholesale (15%) and other (21%). The most popular e-business model adopted was B2B (47%), followed by B2C (18.2%). Some adopted more than two types of trading models. Most e-business development in Australia took place in the year 2000 (22.4%), 2001 (11.2%), 2002 (6.6%), 1998 (16.4%) and 1999 (18.4%). The adoption percentages indicate a downward trend since 2000, which is commensurate with the concept that the dot.com crash slowed or discouraged e-business development. E-business like most new initiatives requires substantial resources in terms of technology, finance, people and time. Most Australian organisations made substantial investment in e-business resources as findings indicate that 71% assigned all development planning duties to a dedicated group, 65% indicated that they allocated large financial resources to e-business projects and 68% had allocated dedicated personnel to manage and implement e-business projects.

E-Business Drivers

E-business drivers according to Barua et al [4] are factors that impact e-business operational and financial performance. In Table 1, we consider the customer orientation of e-business propositions. Thus in the case of the first proposition in Table 1 (customers can contact representatives), there were 62 participants agreeing to the proposition out of a total of 114 responding to this proposition. Here $n = 114$ and $p = 62/114$ or 54.4%. The 95% confidence interval will have a lower endpoint of 45.2% and an upper endpoint of 63.6%. Thus when we consider the responses of small (51.3%), medium (58.5%) and large (55.6%) organisations to this proposition, we see that in each case the response rate is within the confidence interval. Thus we conclude that there is no statistically significant difference between the response levels of each of these groups and the overall response level. In the following tables we mark occasions where the response is below the confidence

interval with a single asterisk (*) indicating that this response rate is significantly below the overall rate and those occasions where the response is above the confidence interval with a double asterisk (**) indicating that this response rate is significantly above the overall rate. The discussion of findings in this paper is concentrated on issues where the difference in values is statistically significant.

Customer orientation data is presented in Table 1.

For customers to interact with business online it is imperative that customers are well oriented with all aspects of e-business. From the findings presented in Table 1, it is apparent that to reach out to customers the large organisations successfully promoted their e-business by listing their URL's on search engines, e-business promotion by medium sized e-businesses is seen to be fairly poor, and small e-businesses did not provide product related information satisfactorily on their web sites. Personalised web pages enable customers to create their own Web pages, and use it to record purchases and preferences, as well as capture information such as product details, and disseminate information on add-on purchases and warranty. This obviously is an advanced technology based feature which is better offered by large organisations only. Although the web technology is useful for establishing online communities with email, bulletin boards, online forums and chat facilities, this feature is seen to be least popular. However, medium enterprises utilised this feature better than others.

In Table 2, we consider the Supplier related processes analysed in the same way as the data in Table 1. For buyers and suppliers to interact in the B2B e-space it is imperative for the suppliers to be well oriented with electronic ordering, accessing and sharing information online with all business partners, participating in online auctions, and supporting the electronic supply chain for the replacement of inventory with information. For supplier interaction, medium sized enterprises were more successful in online business communication and providing information as FAQs as compared to large and small ones.

In Table 3, we consider the E-Business Processes Findings indicate that medium sized organisations developed their processes to serve customers better than large and small organisations

In Table 4, we consider the Readiness propositions. From the findings presented in Table 4 it is apparent that security of data interchange in small organisations is far less developed. It is also clear that this feature is well established in large organisations. The other finding worth noting is that only customers of large organisations are able to interact online with e-businesses.

In Table 5, we consider the System Integration propositions. Systems integration initiatives network the stakeholders (customers, suppliers and internal) systems in an e-business environment. Automated

processes for quick retrieval and processing of information, one set of data or integrated databases, easy access of information and in some cases a single contact point for customers, suppliers and employees are key attributes of e-business that are achieved from integrated system. Automated and integrated business processes form the back end e-business systems. This feature of e-business is significantly well developed in large organisations as evident from the data in Table 5.

In Table 6, we consider the Internal Orientation of IT application. The impacts of intranets and internal communication, project management, internal process improvements and internal orientation of electronic business processes are pivotal for e-business success. From Table 6 it is clear that this feature of e-business was only moderately achieved by all organisations investigated. As stated by Barua et al (2001), it is important for firms to ensure that employees can manage online customers and suppliers with easy to use internal information systems.

In Table 7 we consider the improvement in Financial Performance. The percentage of overall respondents who indicated an above average financial improvement is shown. For example, 32.5% of respondents overall indicated an above average improvement in return on investment, 28.9% on return on assets, 28.1% on gross profit margin and 20.4% on revenue per employee. In Table 8 we consider the improvements achieved from the implementation of e-business. The percentage of overall respondents who indicated an above average improvement in each measure is shown. For example, 55.3% of respondents overall indicated achieving an above average improvement in providing increased information to their customers via e-business. From the data presented in Table 8 the figures portray what e-business organisations have achieved from e-business. Availability of information which usually is a first step in e-business development has been achieved while other aspects of e-business are far from developed.

5. DISCUSSION AND CONCLUSION

Although all of the findings have not been explicated in detail in this paper the findings clearly indicate that e-business is the new way of doing business and adopted widely by Australian organisations irrespective of their size. From this it is inferred that the Internet is increasingly transforming traditional businesses to e-businesses. Although most industry sectors have embraced e-business, manufacturing and service industries are capitalising on the opportunities more than others. It is clear that IT managers popularly undertake e-business leadership indicating that IT knowledge is essential for e-business project leaders. Although it is not new that larger organisations are in a better position to allocate specific resources for new projects, e-business implementation in Australia shows that the medium and small organisations have also invested substantially in e-business. This indicates that the value of e-business has

been realised by all Australian organisations although real benefits of e-business have not been realised. While e-business has obtained some negative publicity in the recent past due to dot.com crash, it is apparent from this research that the Internet is increasingly transforming traditional businesses into e-businesses. This research is one of the first in Australia that has attempted to evaluate the performance of e-business. From the findings presented in this paper, it is apparent that e-business in Australia is not too different from the findings of the US study discussed earlier. We will discuss the value of e-business as it compares to the value in the USA in another forthcoming paper. From the findings discussed in this paper it can thus be concluded that in Australia e-business is very much alive and applied although its implementation is slow and improvised.

REFERENCES

- [1] Australian Bureau of Statistics Report on Household Use of Information Technology 8146.0
- [2] Amit, R., and Zott, C. (2001), 'Value creation in e-business', *Strategic Management Journal*, 22, 493-520.
- [3] Ashburton, C. and Doherty, N. F., 2003, 'Towards the Formulation of a "Best Practice" Framework for Benefits Realisation in IT Projects', *Electronic Journal of Information Systems Evaluation*, Volume 6, Issue 2, pp 1 –10.
- [4] Barua, A., Konana, P., Whinston, A. B., and Yin, F. (2001), 'Driving e-business excellence', *MIT Sloan Management Review*, fall, 36-44.
- [5] Byrne, J., and Kabaila, P., (2002). A new Binomial confidence interval, School of Business Information Technology Working Paper 2. RMIT University, Melbourne.
- [6] Chaffey, D., 2004, E-Business and E-Commerce Management Second Edition, Prentice Hall, UK.
- [7] Cronholm, S. and Goldkuhl, G., (2003), 'Strategies for Information Systems Evaluation – Six Generic Types', *Electronic Journal of Information Systems Evaluation*, Volume 6, Issue 2, pp 65 –74.
- [8] Devaraj, S. and Kohli, R. (2003). Measuring information technology payoff: a meta-analysis of structural variables in firm-level empirical research. *Information Systems Research*, 14 (2), 127-145.
- [9] Dubliss, S., (2000), Retailing and the Internet, proceedings of the First World Congress on the Management of Electronic Commerce, McMaster University, Hamilton, Canada (CD ROM).
- [10] Dunt, E. S. and Harper, I. R. (2002), "E-Commerce and the Australian Economy", *The Economic Record*, Vol. 78 No. 242, pp. 327-342.
- [11] Grey, W., Katircioglu, K., Bagchi, S., Shi, D., Gallego, G., Seybold, D., and Stefanis, S., (2003). An analytic approach for quantifying the value of e-business initiatives. *IBM Systems Journal*, 42 (3), 484-497.
- [12] Kearney, A T, 2002, 'E-Business Outlook: 2003 ATKearney an EDS Company, August.
- [13] Kohli, R., Sherer, S.A and Baron, A. (2003). Editorial – IT investment payoff in e-business environments: research issues. *Information Systems Frontiers*, 5 (3), 239-247.
- [14] Kohli, R., Sherer, S.A and Baron, A. (2003). Complementary investment in change management and IT investment payoff. *Information Systems Frontiers*, 5 (3), 321-333.
- [15] NOIE, 2001, NOIE Report (2001, 'B2B E-Commerce: Capturing Value Online', Commonwealth of Australia
<http://www.noie.gov.au/publications/NOIE/B2B/start.htm> March 2003
- [16] NOIE, 2002, 'The Current State of Play, Australia's Scorecard', www.noie.gov.au
- [17] NOIE Report (2003), The Benefits of Doing Business Electronically - eBusiness.
- [18] Shi, D., and Daniels, L. (2003). A survey of manufacturing flexibility: implications for e-business flexibility. *IBM Systems Journal*, 42 (3), 414-427.
- [19] Singh, M. (2000), 'Electronic Commerce in Australia: Opportunities and Factors Critical for Success', proceedings of the 1st World Congress on the Management of Electronic Commerce (CD ROM), January 19 –21, Hamilton, Ontario, Canada.
- [20] Singh, M., 2002, 'E-Services and their role in B2C E-commerce', *Journal of Management Service Quality*, vol. 12, no. 6, November, pp 434 - 446.
- [21] Singh, M., 2002, 'Electronic Commerce Opportunities, Challenges and Organisational Issues for Australian SME's', *Managing Information Technology in Small Businesses: Challenges and Solutions* edited by Burgess, S., IDEA GROUP PUBLISHING, Hershey, USA, pp 322 –339.
- [22] Singh, M. and Thomson, D., (2002), 'An E-Procurement Model For B2B Exchanges: An Australian Example', 15th Bled Conference on eReality: Constructing the eEconomy, June 17 –19, Slovenia, CD ROM.
- [23] The Age, June, 2001
- [24] Turban, E., King, D., Lee, J. and Viehland, D., 2004, *Electronic Commerce A managerial Perspective*, Prentice Hall, New Jersey.

Table 1: Customer Orientation of E-business

	Overall	Small	Medium	Large
Customers can conveniently contact service representatives or seek service on-line.	54.4%	51.3%	58.5%	55.6%
A comprehensive Frequently Asked Questions (FAQ) section on our company, our products/services, contact information etc is available online.	53.4%	59.0%	51.2%	59.3%
The URL for the company's website is listed on numerous search engines to ensure that customers are able to find the site quickly and easily.	53.2%	52.6%	41.5% *	76.9% **
All product-related information available online (eg price, product description, catalogue, etc).	52.6%	41.0% *	58.1%	46.2%
Customers can customise their orders on-line without phone/fax or face-to-face interactions.	40.2%	32.4%	47.6%	46.2%
Customers can see personalised content (eg products, prices, order history, order status etc) when they log onto the website.	38.6%	31.6%	42.9%	48.1% **
Customers can interact with others using online forums and/or communities (eg bulletin boards, chat rooms, etc).	10.6%	10.3%	24.4% **	3.8%

Table 2: Percentage Agreeing With The Supplier Issues

	Overall	Small	Medium	Large
Changes to order product design are communicated to relevant trading partners electronically in real-time	29.0%	29.4%	41.0% **	14.3% *
Customer feedback and field incidence reports are shared with suppliers/vendors in real-time.	26.5%	25.0%	30.8%	28.6%
Continuously updated inventory information is shared with trading partners online.	25.8%	29.4%	34.3%	19.0%
Continuously updated product demand information (actual and forecasted) is shared online with suppliers	24.5%	25.0%	33.3%	14.3%
Quality information (eg yields) is shared electronically with relevant trading partners in real-time.	22.8%	25.0%	29.4%	15.0%
A comprehensive Frequently Asked Questions (FAQ) section is available online to trading partners.	22.1%	17.6%	32.5% **	17.4%
Frequently updated supplier/vendor evaluation reports are available online.	19.4%	34.4%	26.8%	17.4%
Continuously updated production schedules and capacity information are shared with trading partners online.	16.5%	13.3%	23.7%	13.6%
Online communication (eg discussion forums, online chat) is provided to suppliers.	7.0%	12.1%	10.3%	4.5%

Table 3: Percentage Agreeing with E-Business Processes

	Overall	Small	Medium	Large
Customers generally have one contact point for all service needs.	51.0%	48.6%	67.6% **	40.0% *
There are only a few steps in resolving customer complaints.	48.6%	48.6%	59.5% **	42.3%
Supplier quality monitoring processes are well defined.	46.5%	42.4%	50.0%	56.0%
Standard operating procedures cover all procurement scenarios (eg well defined rules for large versus small procurement).	35.1%	29.4%	42.4%	42.9%
Supplier selection criteria are well documented.	30.8%	25.8%	33.3%	40.0%
Supplier evaluation metrics are well documented.	29.2%	34.4%	35.5%	40.0% **
Information exchange policies with our suppliers (eg frequency, precision, real-time or delayed, format, channel) are well defined.	22.9%	24.2%	23.5%	22.7%
The method for sharing product roadmap and demand forecast with our suppliers is well defined.	22.3%	15.2%	29.4%	30.0%

Table 4: Percentage Agreeing With The Readiness Propositions.

	Overall	Small	Medium	Large
Our suppliers consider it important to improve online coordination and collaboration with us.	46.4%	53.1%	44.1%	40.0%
Customers feel comfortable regarding security and privacy in electronic commerce with our company.	46.0%	35.3% *	45.5%	61.5% **
Our customers consider it important to engage in electronic commerce.	43.6%	41.2%	41.2%	53.8% **
Our suppliers are willing to share information electronically with us.	34.3%	33.3%	36.8%	30.4%
Our suppliers consider it important to engage in e-business.	33.0%	27.5%	25.7%	34.8%
Our suppliers feel comfortable (regarding security, privacy etc) engaging in e-business.	28.4%	21.9%	31.4%	33.3%
Our suppliers have Internet-based systems to engage in e- business.	26.5%	32.3%	26.3%	22.7%

Table 5: Percentage Agreeing With The System Integration

	Overall	Small	Medium	Large
Data can be shared easily among various internal systems in our organisation.	54.5%	50.0%	58.5%	57.7%
Order changes automatically get reflected in downstream processes or systems.	44.3%	40.5%	51.4%	43.5%
Our systems easily transmit, integrate and process data from suppliers/vendors and customers through the Internet.	29.7%	30.3%	27.0%	29.2%
Our systems allow continuous monitoring of order status at various stages in the process (eg manufacturing, shipping).	36.9%	34.3%	36.8%	39.1%
Employees can easily retrieve information from various databases for decision support (eg cost information, reporting tools).	49.5%	42.1%	50.0%	61.5% **

Table 6: Percentage Agreeing With The Internal Orientation Of IT Applications

	Overall	Small	Medium	Large
Employees can retrieve required information through the corporate intranet.	51.4%	52.8%	50.0%	53.8%
Employees manage administrative processes (scheduling meetings, travel, training etc) online.	39.6%	36.1%	41.0%	41.7%
Best practice/project management information is available through the corporate intranet.	33.6%	28.6%	32.5%	36.0%

Table 7: Percentage Indicating A Higher Impact Of E-Business On Financial Performance

Return on investment	32.5%
Return on assets	28.9%
Gross profit margin	28.1%
Revenue per employee	20.4%

Table 8: Percentage Indicating Higher Level of Improvement by Implementing E-Business

Increased information available to customers	55.3%
Increased information available to the company	50.4%
Increased information available to employees	49.6%
Improved customer service	44.3%
Reduction in total costs	29.6%
Retention of online customers	28.9%
Increased information available to suppliers	23.7%
Increase in revenue	22.6%
Reduction in labour	21.1%
Increase in the number of new customers	19.3%
New markets	18.6%
Increase in the number of customers overseas	18.4%
Reduction in inventory	16.7%

E-Business: Factors of EAI Adoption in an Australian Electricity Company

Ying Chen, Alex Wong

School of Information Systems, Faculty of Commerce, University of Tasmania, Australia
Ying.chen@utas.edu.au; Adwong@utas.edu.au

ABSTRACT

The awareness of the scarcity of literature in EAI adoption in the public sector led to the current study of an investigation into the factors that influenced the adoption of EAI in an Australian electricity company. Using a case study approach and qualitative methods, data was collected through semi-structured interviews with five informants from different aspects of the EAI project. Data analysis was conducted using the grounded theory, resulting in 24 themes, which were then categorized into three broad themes of environmental influence, current problems and perceived benefits. These themes were taken as factors that contributed to the decision making of the EAI project of the case company. The informants considered environmental change as the trigger for the adjustment of the case organization's strategy. However, they all shared the view that the adoption of EAI was the consequence of a clear identification of the strategic positioning of the organization, the understanding of existing inhibitors and the awareness of the benefits of a chosen technology.

Keywords: EAI, enterprise systems, decision process

1. INTRODUCTION

Enterprise application integration (EAI) has over the past few years quickly become an alternative option other than Enterprise Resource Planning (ERP) to address organizations' problems on enterprise integration. An obvious indication is the rapid development of middleware technology originally designed for data-oriented integration that has quickly transformed into more sophisticated software architecture.

An examination of existing EAI literature reveals more research has been conducted in relation to the private sector as compared to the public sector. A possible explanation is that the public sector has been less dependant on IT than their private counterparts that utilize IT heavily to greater profitability. The bureaucratic structure of the public sector as well as government policy and regulation controlling competition has resulted in a more conservative approach to IT adoption [15]. However, the operational environment is changing for the public sector. In Australia, the government relinquished control of public sector investments in the 1990s, which included telecommunications and energy distribution industries, an endeavour known as deregulation. Deregulation transformed the operational environment of the public sector from a stable and predictable market to one characterised by dynamism and complexity, in which technology became a valued commodity [15]. The public sector became government business enterprises (GBE) and is required to generate profit.

The research presented in this paper aims to investigate an EAI project in a GBE within the electricity industry in Australia after deregulation. To confront competition with other electricity providers within Australia, the

company decided that it required an enterprise system that should streamline the data flow between three major business sectors within the organization: the human resources department, the assets department and the accounting department. The organization examined two alternative approaches of enterprise integration, ERP from one vendor and EAI with the best of breed. It was decided to pursue the latter option. This study intends to investigate the factors that influenced the decision to implement an EAI project. The findings will be of value to researchers in the area of EAI and the public sector and to organizations with a similar endeavour.

2. LITERATURE REVIEW

2.1 Importance of EAI

EAI is an umbrella term that encompasses technologies designed to unite disparate applications to address business problems. It aims at the unrestricted sharing of data and business processes among any connected applications and data sources in an enterprise [11].

Information technology has been utilised in organisations as a means of automating and improving business processes. However, before 1990s, IT was used to address problems based upon the unique needs and beliefs of business units within an organization, resulting in an IT architecture characterised by autonomous and heterogenous solutions [11]. This resulted in inefficient processes, reduced quality and accuracy of information that potentially impacted on coordination of activities and operational efficiency [8].

With increasing competition in the modern business environment, organisations view knowledge as the only source of competitive advantage with which they may

differentiate themselves from competitors [3]. The inherent decentralized IT architecture creates a big challenge for organizations' strategic planning and decision-making [12].

The trend towards E-business, E-commerce, corporate mergers, acquisitions and globalization in the past few decades has also created other incentives for an integrated enterprise system [2].

ERP emerged in the early 1990s as an enterprise-wide solution to address organisation's coordination issues. In 2001, the enterprise system industry was estimated to be worth US\$19 billion [5]. However, the complexity and cost of ERP in combination with the current economic crisis has made ERP an unattractive option [18].

EAI surged as an attractive alternative option for reduced cost [4, 21], for maximising an organization's existing investments as well as maintaining a unique business process that differentiates an organization and gives it a competitive edge [10]. The cost reduction of EAI was also evidently important to public enterprises [7, 8] as they do not have the high profit margins of their private sector counterparts.

2.2 Challenges of EAI

Challenges of EAI range from the incompatibility of different software, lack of standards in IT architecture and lack of expertise [11]. Middleware technology is a fast developing area aiming at the rising EAI market, but the currency of the technology sees a relative scarcity of individuals with the proficient skills necessary for application integration projects. Despite the fact that previous researchers have found EAI a cheaper option than ERP [4, 21], a full EAI implementation can cost up to AU\$10 million dollars, and in many cases represents the largest IT expenditure in the history of an organisation [20]

2.3 Public Enterprise and EAI

The public sector differs from its private counterpart with some unique attributes. It enjoys a significant establishment time, sometimes over 100 years [2], thus has cultivated the perfect breeding ground for a complicated, inconsistent software architecture [2]. It has a large number of employees requiring extensive management and coordination [7, 8]. It is geographically dispersed as a result of acquisitions and mergers to cut down overhead costs [7]. The culture of the public sector displays signs of bureaucratic control and homogeneity, resulting in a slow response to environmental changes. However, this culture will have to change due to deregulation that brings with it consistent unanticipated changes [2].

Although EAI is viewed by some researchers as a better fit for the public sector [7, 8, 12, 17] research within this context is limited. The existing literature reveals EAI has been adopted by public enterprises such as healthcare, education and telecommunications [7, 8, 21]. In the context of the electricity industry, researchers have identified two triggers of EAI. One relates to the wave of mergers and acquisitions and the other to the sudden increase in the adoption of IT as a result of deregulation. [2, 6, 14].

3. RESEARCH METHOD

The research question of the study is "What factors influence the adoption of EAI within an Australian GBE in the electricity industry?" The research, based on a case study technique, is subjective in nature and takes an interpretivist stance to explore informants' personal experiences, perceptions and opinions in relation to the EAI project within the organization.

The interview protocol, with a detailed design of questions to probe for information most relevant to the research, was pilot tested by the IS manager of the project for flow, timing, content and comprehensibility. Based on the feedback from the pilot test, the interview protocol was refined.

Five informants, representing various roles in the steering committee of the EAI project, were invited to take part in the research. Semi-structured interviews, each approximately one hour in length, were tape-recorded with the consent of the informants and transcribed in accordance with Yin's 24-hour rule [22]. To address validity, the research recorded in detail all aspects of the research process. Both the transcripts of the interviews and a copy of the findings of the research were provided to the informants for validity checks. Close collaboration with the IS manager of the project was conducted in relation to the identification of informants, interview design, accessing internal documents and clarification of issues and terms unfamiliar to the researchers. This practice was in close alignment with the recommendation of three methods for achieving validity from Bailey [1]: natural history, member validation and collaboration.

The research addressed the issue of reliability through a complete transcription of the interviews and a detailed documentation of the coding process to reach the final conclusion. This effort recommended by earlier researchers [1, 22] was to facilitate replication of the study by other researchers.

The data were analysed using a bottom-up, three-stage coding process: open, axial and selective as devised by Strauss and Corbin [19]. Each 'pass-over' was intended to refine and condense the data by identifying salient themes. Dominant themes were expected to be explicit by the conclusion of the coding process.

4. FINDINGS

In order to facilitate discussion, and for ethical reasons, the informants are represented with their positions, as shown in Table 1: Profile of Informants. The initial coding of the interview data generated 24 unique themes, that were then categorized into three broad themes, namely environmental influences, perceptions of current problems and perceived benefits as shown in Table 2: Categorization of Emerging Themes.

4.1 Environment Influences on EAI Adoption

An analysis of the data indicated that a series of external factors contributed to the organizational decision on EAI adoption. Most informants acknowledged that the push to enter the National Electricity Market changed the organization’s operational environment as some informants stated:

*We’re going from a fairly sedate market into a national commercial environment (External Contractor).
The entry to a national market essentially exposes us to full competition with other providers (IS Manager).*

Table 1: Profile of Informants

Position	Project Involvement	Employment in Company
IS Manager	Steering committee member	2 years
Process & operations (PO) Manager	Experience with similar projects	15 years
Financial Systems Manager	Consultative committee member	6 years
External Contractor	Project Manager	2 years
Engineering Manager	Business client	2 years

Table 2: Categorization of Emerging Themes

Broad Themes	Underlying Themes
Environmental Influences	Entrance to market; Competition; Competitive positioning; Opportunities; Shareholder ROI; Evolving business
Current Problems	Lack integration; Geographic dispersion; Low assets knowledge; Slow decision-making; Low IT expertise; Change resistant culture
Perceived Benefits	Efficiency; Data accuracy; Information accessibility; improved decision-making; improved productivity; cost control; investment maximisation; standardisation; Extensibility; Systems usability; High ROI; Risk reduction

New external products such as natural gas triggered the need to compete and define the organization’s competitive positioning.

*It’s more important that we position ourselves correctly in the market (IS Manager).
The performance of the asset is an important contributor to being part of the competitive market (External Contractor).*

When we go into the national market, we’re supposed to have machines that have high availability (Financial Systems Manager).

However, a perception of new opportunities created another environmental factor.

*If we can improve capacity a minute portion, that provides sales opportunities (External Contractor).
The entry to a competitive market itself raises the opportunity to exploit higher availability (of assets). We could end up managing assets interstate – which would create another level of complexity... The degree to which opportunities could be exploited, were dependent upon the performance of the assets...essence of the project (IS Manager).*

Other environmental factors included shareholders’ interests and the need to evolve the business.

*The major stakeholder, the government, will know we are running these things, we are sort of maximising the dividend and the tax that we have to pay to them (Financial Manager).
It is quite conceivable that within 4-5 years that 20 –30% of our energy will be derived from wind – and that’s a big change and the new EAI system is expected to manage the maintenance of the assets of wind farms (IS Manager).*

4.2 Current Problems Leading to EAI Adoption

A series of problems were identified that initiated the necessity of the EAI project for the case organization. The first is the awareness of “no integrated approach to anything” as the External Contractor observed.

*We could see a big gap in the organisation in terms of the (assets) maintenance management system (that) was really stand-alone (Financial Manager).
We just need something to pull it all together. It’s like having many horses. They all pull in different directions, the cart, not one coach driver to pull the thing into the right direction (Engineering Manager).*

The second problem lies in the geographically dispersed nature of the organisation.

*Geographic dispersion works against us in terms of being able to have very effective process and good maintenance procedures’ (IS Manager).
Geographic dispersion has caused massive differentiation between processes; local processes rather than uniform processes across the entire company (Engineering Manager).*

Geographic dispersion appeared to be a particular problem for process standardisation attempts. It suggested that the organisation’s control over geographically dispersed factions was tenuous.

The third identified problem was the lack of sufficient understanding and knowledge concerning their assets.

The second, I guess, main aspect of what they need to do is that employees need to improve their understanding and management of those assets (IS Manager).

How long did it take to fix (the problem with assets)? Well the current answer is: 'it depends.' Well it shouldn't depend. Breakdowns of a certain calibre should be fixed in a certain period of time (Engineering Manager).

The informants felt the lack of systems integration and geographic dispersion resulted in a slow organisational decision-making.

Decisions tend to be taken in a fairly measured way and that can be a frustration sometimes because... you feel you need to respond quickly to something (IS Manager). We have a certain number of hurdles you have to go through to reach a decision (Financial Manager).

The informants also pointed out the low level of IT expertise within the organisation.

We need some good IT professionals as well as internal people (External Contractor). There would be a fair scarcity of expertise. The organisation's fairly, you know, short at the moment as far as those expertise (PO Manager).

Further probing revealed that the IT component of the organisation had been outsourced to a third party. The last identified problem related to the organization's resistance to change.

I think in general, our acceptance and willingness to change could be a lot better... I think more people resist change... People are just not prepared to really embrace a case for change (Financial Manager). Some people... don't have a lot of contact with the corporate side of the business and they've been there for a long time and are pretty entrenched in the way they do things, so sometimes, trying to get people like that to change is difficult (External Contractor).

4.3 Perceived Benefits of EAI Adoption

Analysis of the data indicated that informants were well aware of the benefits that the EAI project would bring to the organization.

The first benefit is greater process efficiency as a result of the increased speed of information retrieval, the accessibility, and improved accuracy and quality of data.

It will streamline a lot of the tedium they have to put up with (External Contractor). There will be asset efficiencies, and people utilisation by us having the information at our fingertips rather than having to search for it, make phone calls et cetera... It is supposed to provide them with the right information at the right time, it is supposed to save them time by not having to put information into that system... (Engineering Manager). At the end of the day the organisation should hopefully finish up with better data. They (steering committee) are after up-to-date, complete data on their generation assets, and I don't believe they're getting that at the moment (Financial Systems Manager).

The informants felt an integrated system with improved accessibility and quality of information would also result in better operational decision-making in asset maintenance. Its significant role in this assets-based organization can be seen with these statements.

Generation management are able to make better calls about when (to) and when not to run the machine. They'll have a lot better ability to make some of these decisions once they have an integrated system (Financial Systems Manager). A fully functional platform, which gives me the right information when I need it to make a right decision on the spot. (Engineering Manager).

It was believed that an integrated system would improve employee productivity and reduce operation costs.

Instead of spending our time doing time sheets three times a day, we only do them once, and in the two hours that we save, we're actually doing something a bit more useful towards managing the assets... It's certainly reducing a lot of our operating costs (External Contractor).

All informants indicated that the reduction of cost could be realized through efficient assets maintenance management, mainly the maximization of the life span of existing assets. A rippled effect would be a higher return of investment (ROI).

We can predict with reasonable certainty that the life of this particular piece of equipment in service is so long. So we will aim to maintain it inside that window (IS Manager). It (the integrated system) is to provide us with a suitable platform to maximise the return on investment, these assets which is the essence of asset management – get the best bang for your buck – this is the ultimate objective (Engineering Manager).

Standardisation of processes for geographically dispersed organization was another crucial benefit.

There is an impact there or a value in having a system that brings commonality to the way these people work (IS Manager). If I put this system in, all the regions are going to adhere to exactly the same processes, and you don't get a 'cowboy' manager in one area not adhering to it (Financial Systems Manager).

The integrated system with its uniform processes and improved information sharing would ensure better asset management, thus reducing the level of risk toward the organisation.

Unless they (assets) are maintained (with) a regular schedule (and) kept up to date the whole process of when, where, and how (they are) monitored - we put the whole organisation at a fairly big risk (PO Manager). If you start trading, signing contracts, and then your asset doesn't work to fulfil those contracts, you're likely to be hit with a fairly severe penalty (External Contractor). If we don't get our operation right, obviously we have the potential to be undercut, lose market share, lose revenue, potentially affecting profit (IS Manager).

Other benefits included the new system's potential to support new emerging products and services of the organization.

I imagine once (the system is) implemented, they'd (high level decision makers) probably see other advances, they'd say "OK, you can do this", they'd probably look at another area and do this as well, so I'd imagine it would get bigger or have more required of it as time goes by (PO Manager). I've got an idea because of its nature it will evolve gradually over time (Financial Systems Manager).

The informants also believed that the system would support high usability and generate high returns.

Usability would be high (PO Manager). They're also really strong on wanting a usable system (External Contractor). The payback period is only about eighteen months. It has a very high return on investment that is quickly paid back (IS Manager). The internal rate of return of the project when we put everything in there it came out as something 150% of 200% something huge – that's the beauty about such maintenance improvement, that anything that can assist the plant performance, directly or indirectly, has got a huge payoff (Engineering Manager).

5. DISCUSSIONS

The data analysis shows that there are three contributing factors to the organizational decision on the EAI project: a thorough understanding of the operational environment of the organization, a clear identification of existing problems and a good knowledge of the benefits of the EAI project to the organization. The relationships of the three elements are demonstrated in Figure 1.

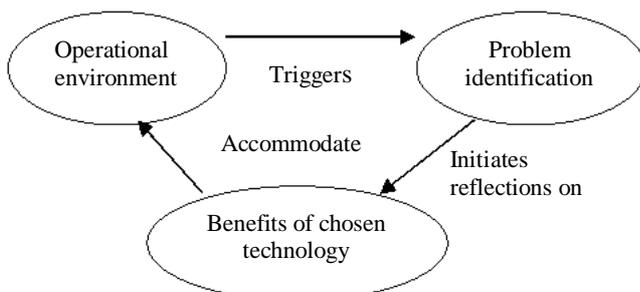


Figure 1: Relationship of factors of EAI adoption

This decision-making process is in alignment with the organismic approach to decision theory in which environment plays a significant role [9]. The organizational practice also suggests a close affinity with the game theory [9], which incorporates both environmental influences and organisational objective as being responsible for an organisation's actions. Another interesting observation was the similarity between the case organisation's objectives and those of private organisations. Foer and Moss [6] observed that private enterprises have an "entrepreneurial culture" that is characterised by an undue focus on profitability, reduced product cost, an increase in product diversity,

and meeting customer's needs. The case organization has adopted most of them except product price. This suggested that deregulation, to a certain degree, had served to dilute the characteristics that differentiated the public and private sectors. However, the public sector might still see an edge in their product availability rather than reduced prices. A possible explanation is a public organization is usually heavily assets based that gives it the potential of providing better service if the assets are well managed and utilized for maximum productivity.

The geographically dispersed nature of the company had its merits in pre-deregulation days when electricity generation companies were responsible for providing electricity to vast land masses [13] and decentralization provided a degree of flexibility [15]. However, in the deregulated regime, geographic dispersion became a weakness when competing against centralised rivals [13]. Furthermore, decentralisation was also viewed as a weakness for public sector organisations as it results in higher operating costs as compared to centralised rivals [16]. The informants expressed similar concerns, blaming the geographic dispersion to be the cause of inconsistent processes that had reduced the organisation's ability to perform efficiently.

The organisation's slow decision-making process can be viewed in the context of lack of information due to dis-integrated systems and the culture of public sector organisations. As mentioned previously, the culture of many public sector organisations tended to reflect their monopolistic nature [6]. Aspects of the monopolistic nature included a bureaucratic structure, and aversion to risk [6]. This implies that public organizations have a culture more unfavourable for change. Any business process reengineering endeavour that involves new cross-functional business processes can be expected to incur more resistance. This is certainly supported by the interview data of the study. The 'Mateship' program used in the organization to encourage loyalty of its employees seemed to further complicate any change initiatives. This accounts for the critical role of the "usability" as identified by the informants. Research indicates an EAI system is a more attractive solution for its minimal impact on culture change, which may account for the case organization's decision to take EAI instead of ERP.

The low level of IT expertise is common within public sector industries. However, like other public organizations that chose to resort to IT for competition [14], the subject organisation used the same strategy and attempted to enhance and cultivate its IT skills through training and recruitment.

The informants' perceptions of EAI benefits are consistent with existing literature. The case organization's paramount interest in work efficiency reflects the observation of many former researchers [17,

7, 21, 2, 8]. Also, the data analysis reveals the inter-relationship between the perceived benefits of EAI projects. Process efficiency was to be achieved through the increased accuracy and accessibility of data, that yields numerous other benefits such as improved organisation's decision-making process, increased productivity of employees and a better control over both cost and risk in relation to assets maintenance. The linked system, as a result of the EAI project aims to help overcome the physical boundaries, a big issue among the majority of public organizations, and help facilitate greater communication and greater control of a standardised business process.

6. LIMITATIONS AND FUTURE RESEARCH

As this research takes an interpretivist stance, the researchers of this study consider it improper and unrealistic to generalise the results to other organizations. However, the research provides some insights into the decision process of adopting advanced technology for the public sector, which will be of value to both academic research and practitioners. A number of issues have emerged for possible further research. They include the impact of organisational culture or top management support on EAI adoption within the public sector and the degree to which the objectives and operations of the public and private sector converge after deregulation.

REFERENCES

- [1] Benbasat, I., Goldstein, D. and Mead, M. (1987) 'The Case Study Research Strategy in Studies of Information Systems.' *MIS Quarterly*, 11 (3): 369 – 385.
- [2] Chang, K., Gold, A. and Kettinger, W. (2003) 'The Extent of Enterprise System Adoption in Companies: A Multiple Theoretical Perspective.' Florida, *Ninth Americas Conference on Information Systems*, August, 2003.
- [3] Davenport, T. (1998) 'Putting the Enterprise into the Enterprise System.' *Harvard Business Review*, July-August, pp122 – 131.
- [4] Demetriades, J. (2000) 'Twenty Questions for Jim Demetriades.' *EAI Journal*, January, pp. 44 – 47.
- [5] Fan, M., Stallaert, J. and Whinston, A. (2000) 'The Adoption and Design Methodologies of Component-based Enterprise Systems.' *European Journal of Information Systems*, 9 (1): 25 – 35.
- [6] Foer, A. and Moss, D. (2003) 'Electricity in Transition: Implications for Regulation and Antitrust.' *Energy Law Journal*, 24 (1): 89 – 100.
- [7] Grimson, J., Grimson, W. and Hasselbring, W. (2000) 'The SI Challenge in Healthcare.' *Communications of the ACM*, 43 (6): 48 – 55.
- [8] Khoubati, K., Themistocleous, M. and Irani, Z. (2003) 'Conceptual Model for the Adoption of Enterprise Application Integration in Healthcare Organisations.' *Ninth Americas Conference on Information Systems*.
- [22] Kmietowicz, Z. and Pearman, A. (1981) *Decision Theory and Incomplete Knowledge*. Aldershot: Gower.
- [10] Lee, J., Siau, K. and Hong, S. (2003) 'Enterprise Integration with ERP and EAI.' *Communications of the ACM*, 46 (2): 54 – 60.
- [11] Linthicum, D. (2001) *Enterprise Application Integration*. London: Addison-Wesley.
- [12] Markus, M and Tanis, C. (2000) 'The Enterprise System Experience – From Adoption to Success.' In Zmud, R. (ed) *Framing the Domains of IT Research: Glimpsing the Future Through the Past*. Cincinnati Ohio: Pinaflex Educational Resources, pp173 – 207.
- [13] Navarro, P. and Shames, M. (2003) 'Electricity Deregulation: Lessons Learned from California.' *Energy Law Journal*, 24 (1): 33 – 55.
- [14] Nehrir, H., Gerez, V. and Holland, S. (2002) 'Electric Power Generation and Management.' *Montana Business Quarterly*, 40 (3): 18 – 23.
- [15] Robbins, S., Bergman, R, and Stagg, I. (1997) *Management*. New Jersey: Prentice-Hall
- [16] Shaw, M. (2000) 'Building an E-Business from Enterprise Systems.' *Information Systems Frontiers*, 2 (1): 7 – 17.
- [17] Souder, W., Daniel, S. and Cooper, D. (2298) 'Environmental Uncertainty, Organisational Integration, and New Product Development Effectiveness: A Test of Contingency Theory.' *Journal of Product Innovation Management*, November, 520 – 533.
- [18] Stonebraker, M. (2002) 'Too Much Middleware.' *ACM SIGMOD Record*, 31 (1): 97 – 106.
- [19] Strauss, A. and Corbin, J. (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. California: SAGE.
- [20] Sumner, M. (2000) 'Risk Factors in Enterprise Wide Information Management Systems Projects.' *Proceedings of the 2000 ACM SIGCPR Conference on Computer Personnel research*.
- [21] Themistocleous, M., Irani, Z. and Shariff, A. (2002) 'Evaluating Application Integration.' *Proceedings of 7th European Conference on Evolution of Information Technology*, pp. 193 – 202.
- [22] Yin, R. (1994) *Case Study Research: Design and Methods*. California: SAGE.

Electronic Business in Germany: Current Challenges and Future Perspectives - Results of an Explorative Survey -

Tomaso Forzi, Ediz Kiratli, Stefan Bleck

E-Business Engineering Department of the Research Institute for Operations Management (FIR)
at Aachen University of Technology (RWTH Aachen), Pontdriesch 14/16, 52062 Aachen, Germany
{Tomaso.Forzi, Ediz.Kiratli, Stefan.Bleck}@fir.rwth-aachen.de

ABSTRACT

The objective of the present contribution is to present the results of an explorative survey in the field of E-Business conducted with German enterprises at the end of the year 2003. The survey aimed to convey a picture of the present stage as well the future trends of E-Business in the German entrepreneurial world. An analysis of the present and future developments of E-Business as well as of the related organisational, technical and human implications for enterprises will be presented. Limitations and open challenges of research as well as the planned further work will be eventually discussed.

Keywords: Electronic Business, Explorative Survey, Germany

1. BACKGROUND OF RESEARCH

1.1 Introduction

During the past years, the fast development of new Information and Communication Technologies (ICTs) has been revolutionising the market arena, it extended the horizon of competition, and it caused radical changes in all business branches [3] [12]. The effects of the introduction of E-Business (EB) were an enhanced globalisation process, an even more uncertain and dynamic economic environment, and a technology-driven development of new capabilities, products and services as well as new businesses [12] [13]. This was the case also in the German economy and entrepreneurial structure [3] [13].

In order to analyse the current state as well as the implications of the short but intense history of EB in the German entrepreneurial world, as well as in order to identify key success factors, challenges and future trends, at the FIR it was decided to start a series of empirical works, which the present work belongs to.

1.2 Scope of the explorative study

In this contribution we are going to present the results of an explorative survey, entitled "E-Business in Germany –Success Factors and Obstacles", which was conducted with a sample of German enterprises at the end of the year 2003. The survey aimed to convey a picture of the present stage as well the future trends of EB in the German entrepreneurial world. In particular, the research strived to analyse the current state of EB in German enterprises in order to identify its relation with their economic success, strategy, organisation, processes, ICT infrastructure, and personnel. Furthermore, critical success factors as well as most significant obstacles for companies within the deployment of EB were also sought.

This explorative survey has to be seen as the initial step of a wider research framework; as a matter of fact, its ultimate objective is to enable an initial analysis in order to identify patterns and research areas to be then further and deeply analysed within following empirical works.

2. RESEARCH METHODOLOGY

2.1 Phases of research work

The research work was structured in the following phases:

1. *Selection of an appropriate approach of empirical research*, in which after an analysis of different methodologies of market research [6] [7] [8] [9] [10] [11] [14] [15], the approach of explorative survey with partial census and proportional selection of participants was found to be the most appropriate within the several approaches of primary research. In particular, the survey was planned to internet-based [1]; all participants were to be approached only via email and were asked to fill in the questionnaire online. An option to fill in the questionnaire offline and to send it back by fax or snail mail was also to be offered.

2. *Survey design and technical implementation*, in which the core elements of analysis were identified, the different hypotheses were listed and structured, and the questionnaire was hence designed taking into consideration the requirements of empirical research [1] [6] [7] [8] [9] [10] [11] [14] [15]; furthermore, the execution of a pretest with selected companies helped to gain precious feedback that was integrated into questionnaire, which was hence implemented in a web interface and eventually tested.

3. *Sample selection*. Within this phase, the first step was to define the indices according to which the entrepreneurial quota were going defined: industrial sector as well as size of the enterprises [8] [9] [10] [14]. Such indices were obtained from the Classification of

Economic Activities in the European Community (NACE). The quota related to such indices for the proportional selection process of the survey participants were hence identified with the help of the data from the German Federal Statistical Office¹. Hence, the list of enterprises to be invited to take part to the survey was prepared with data from the database of the market research firm Hoppenstedt² and the own FIR database.

4. *Realisation of the survey.* In this 3 month phase the selected enterprises were sent an email with an invitation to take part to the survey as well as some following reminders, also via email. It is interesting to note that 72,6% of the participants filled in the questionnaire online, while 27,4% preferred to fill in the printed questionnaire and send it back with the support of more traditional means (23,4% via fax and 4% via snail mail). This shows that, although all the participants had an internet access, about one third felt more comfortable with the offline and manual processing of the questionnaire.

5. *Follow up telephone interviews.* Since in the paper-based questionnaires some inconsistencies were found, some follow up telephone interviews were needed in order to clear few open issues with some of the participants [11] [14] [15]. This action enabled the clarification of the open issues and enabled the further use of the whole data set within the analysis phase.

6. *Collection and coding of data.* The automatic saving of the data of the online questionnaire in a previously programmed database enabled an efficient and error-free coding of such data; the paper-based questionnaires were manually coded and accurately doublechecked.

7. *Data analysis and documentation of the results.* The collected data were analysed with the help of univariate, bivariate, and multivariate statistical methods [2] [5] [11] [14]. Appropriate statistical testing methods were used to test the results of the analysis: Significance-test with $\alpha < 0,05$ for the tested hypotheses, Chi²-test for bivariate and contingency analysis, F-test within the multivariate analysis. Furthermore, in order to avoid distortions caused by different scales (especially in the multivariate analysis), all different units and scales were standardised with the help of the Z-transformation. Cluster analysis that were conducted with different standardisation methods lead always to the same results. Eventually, all analysis and results were documented.

2.2 The sample

During the sample selection 8.150 enterprises (gross sample) that fulfilled the criteria were selected. Decision-makers of such enterprises were sent an email with the invitation to take part to the survey. During the mailing action 1.258 emails did not reach the targeted group, this meaning that 6.892 potential participants received the invitation email (net sample). By an analysis of the net sample it came out that the striven quota of

population, with an acceptable deviation, were still fulfilled by the net sample.

On the whole, 278 enterprises took part to the survey and sent back the filled questionnaire, with the result of rate of return of 4,03%. The distribution of the participants vs. the values of the population are given as follows: 48,6% vs. 21,4% in the industrial sector, 42,1% vs. 40,1% in the field of services (without commerce); 9,3% vs. 24,7% in the commerce sector, and 0,0% vs. 2,3% in the agricultural sector. As far as the entrepreneurial size is concerned, the distribution of the participants vs. the values of the population are: 62,2% vs. 89,4% small sized enterprises, 17,3% vs. 10,3 medium sized enterprises, 20,3% vs. 0,3% large sized enterprises. Some branches and some enterprise sizes are overrepresented, while other are underrepresented.

Because of the high unit non-response, a non-response analysis was conducted. The major reasons of non-response are lack of reachability, technical limitations, and noncompliance [4]; while the lack of reachability is the most appropriate explanation for the difference between gross and net sample, the technical limitations could be to a certain extent excluded because of the way in which the web questionnaire was implemented; as a result, the unit non-response of the net sample is likely to have its origin in the noncompliance of the invited persons [4]. An analysis of the reasons of non compliance would have required a new survey of a sample of the persons that did not fill the questionnaire in: this new survey would have gone beyond the scope of the present work.

As a matter of fact, since the objective of the empirical work was to conduct an explorative survey, and not necessary a representative survey (planned for a further research phase), the data was considered sufficient for the explorative scope of the research.

2.3 The questionnaire

The questionnaire of the survey had 39 questions divided in seven sections with a total of 556 items. It covered the following areas:

1. *E-Business in the considered company and in the economy*, in which along with the different items about the present state of EB in the own firm (e.g. involved areas, duration, strategic objectives as well as achievements of EB), also a qualitative evaluation of the state of EB in the considered firm, industrial sector and German economy in the present and in the coming one to two years was gathered. Furthermore, the issue of EB related innovation in the considered firm, its competitors and in the related industrial sector were dealt with.

2. *Strategy in E-Business*, in which data about strategic aspects of the EB deployment were collected, such as e.g. the existence of a EB strategy and its relation to the entrepreneurial strategy, experiences in the strategy

¹ see also http://www.destatis.de/e_home.htm

² see also <http://www.hoppenstedt.de/>

finding process as well as experiences with approaches and methods used within such a process.

3. *Organisation and processes in E-Business*, in which organisational and process-related issues (such as the positioning of EB within the organisational hierarchy or the extent of the reorganisation of the entrepreneurial processes) as well as operational objectives of EB were focused on. The extent of reorganisation of processes was also inquired.

4. *Information and Communication Technology in E-Business*, in which EB budget and its temporal development, typology and evaluation of the involved solutions as well as of the involved ICT service providers, the information sources used for within the selection of the ICT service providers and the software criteria that ex-post were found to be critical.

5. *Personnel and E-Business*, in which data about EB trainings (such as type and targeted employees), a qualitative employees' evaluation of the EB deployment within the firm as well as of the employees' productivity variations caused by EB were collected.

6. *Further comments* about Electronic Business, could be then added within a section for free text remarks.

7. *Data about the interviewee and demographic data of the enterprise*, such as NACE-Sector, number of employees, turnover in the previous year as well as age of the firm.

3. RESULTS

3.1 Selected results

In this chapter some of the results of the monivariate analysis of data are going to be presented and discussed.

1. *State-of-the-Art and E-Business in the future*. On average, the enterprises have been dealing with EB for 3 to 4 years and mainly see it as a strategic approach to conquer new market segments and new customers (27,4%), to increase their turnover (17,8%) and to improve their relationships with their customers (17,6%). While currently the degree of EB maturity of nearly half of the enterprises is considerably low (e.g. 48,4% of the firms concentrates their activities on a web presence and web marketing), in the coming future the trend is clearly towards a more complex and manifold understanding of EB deployment (e.g. 19,6% plan a holistic EB concept or even 25,7% an EB supported inter-organisational integration), see also Figure 1. This confirms the hypothesis of a trend towards an increasingly networked economy [12]. As far as the entrepreneurial functions are concerned, the ones which are currently mostly involved within EB are sales/ marketing (76,4%), procurement (61,6%) and after sales service (45%).

In the future, while de facto all firms will operate sales/ marketing over the Internet, almost 80% will be practicing e-procurement and web-based after sales services, about 70 % will be practicing web-based Customer Relationship Management (CRM). As far as

approaches of inter-organisational integration are concerned, within the coming years 60% of enterprises will implement e-logistics solutions, while about 40% solutions for Supplier Relationship Management (SRM) and Supply Chain Management (SCM). This result is consistent with the trend highlighted in Figure 1.

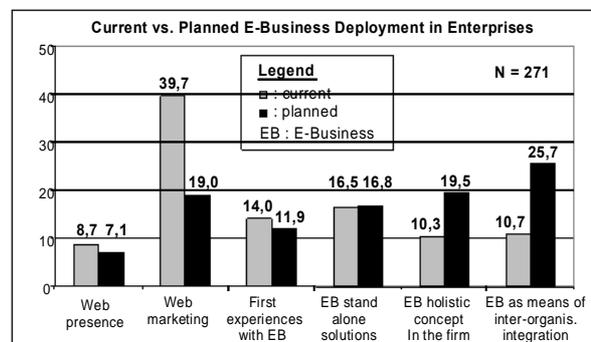


Figure 1: Current and planned EB deployment

While only a part of enterprises could achieve an increased turnover (33, %) or lower procurement cost up to 10% (29,3% of the firms) through EB, the majority of them estimates EB as important or very important for their enterprise (63,6%), while moreover 84,2% declare to plan to increase their EB deployment in the next one or two years. Comparable positive estimations were given regarding the EB growth in the own industrial sector and overall in the German economy in the coming future by 84% to 92% of participants.

2. *E-Business Strategy*. Within the analysis of data about strategic aspects of the EB deployment, it was noted that, even if only 37,8% of the firms formulated an explicit EB strategy, it was almost always adjusted and aligned to the entrepreneurial strategy (91,6% of them). This process is clearly important, especially if we consider the previous experiences related to the E-Business hype and to the hastily embedding of EB-solutions into organisations [12]: firms that have an EB strategy, are developing it with a structured and methodical approach and furthermore in alignment with the entrepreneurial strategy.

3. *Organisation and processes in E-Business*. Within the analysis of the positioning of EB within the organisational hierarchy, no dominant approach could be found. EB is either in the competence field of the management board through a dedicated staff unit (21,6%) or it is given an own organisational unit (only 6% of firms) or it is assigned a third level position in the organisational hierarchy, as subunit of an organisational unit (31,3%), mainly sales/ marketing or ICT. Interesting is that in 28,7% of firms uncoordinated EB competencies are to be found.

The extent of the reorganisation of the entrepreneurial processes following the introduction of EB was also inquired: only 22,8% of the firms carried out a (even partial) reorganisation of processes related to the EB

introduction. This fact seems to be a surprising result and it will be further discussed in section 3.5.

Furthermore, operational EB-objectives were also analysed, with the following top-three-ranking: better customer retention (23,6%), lower process costs (16,3%), and better communication with customers and suppliers (15,6%).

4. Information and Communication Technology in E-Business. Within the analysis of ICT solutions for EB, it came out that only 17% of the firms make use of Application Service Providers (ASP), while 83% implemented the solutions in the own enterprise; furthermore, the vast majority of such solutions are either standard solutions, with (51,0%) or without customisation (19,0%), or fully new solutions developed only for the own enterprise (21,9%). As far as the involved ICT service providers are concerned, they were selected mainly based upon information gained through personal contacts or recommendations, hardly ever after a systematical and structured analysis of capabilities and performances of the specific solution. Nonetheless, both the implemented ICT solutions and the collaboration with ICT service providers were evaluated as good. Furthermore, the software criteria that ex-post were found to be critical were the user friendliness, the capability of system integration, and flexibility of software.

5. Personnel and E-Business. It must be highlighted that enterprises are aware of the fact that the employees play an important role within the successful EB deployment. As a matter of fact, the majority of enterprises gives the own employees a generic training about Internet and EB (89,9% of the firms) and more specific training for the use of EB solutions (77,7%). Still, some differences arise while analysing the profile of the targeted employees.

On the whole, the employees of the participant firms welcomed EB in their daily work; in fact, in 61,1% of the firms, the employees' perception and consequent evaluation of EB is positive to very positive, in 25,7% neutral and only really seldom negative (1,8%). Last but not least, a significant amount of enterprises (40%) underlined that the productivity of the own employees increased because of EB.

3.2 Differences among small, medium and large enterprises

With bivariate and multivariate analysis of the data, differences among enterprises of different enterprise sizes were analysed. On the whole it can be stated that there is a remarkable gap between large enterprises (LEs) on the one side and small enterprises (SEs) on the other side. Medium enterprises (MEs) are, depending on the issue, either leading with LEs or following together with SEs.

As a matter of fact, LEs are, within the degree of deployment of EB, pioneers with a holistic EB-concept (22,9%) or even inter-organisational integration (25%) and are in clear advantage to SEs and MEs, which are by far beyond LEs with e.g. initial EB-experiences (16% and 16,4%) or even simpler web marketing (44,8% and 38,8%) approaches.

As far as strategic issues are concerned, 52,8% of LEs and 45,7% of MEs have defined an EB-strategy, while only 30,8% of SEs defined it. The differences become smaller in the case of a planned EB-strategy in the future (69,7% LEs, 66,7% MEs, 66,7% SEs), this meaning that, even if coming from different experiences, enterprises are increasingly aware of the importance of the definition of an EB-strategy.

As far as strategic EB-objectives are concerned, LEs strive mainly to EB-driven cost reductions, while SMEs strive more to acquiring new customers. This aspect is confirmed by the operational EB-objectives, according to which LEs strive first of all to lower process costs (a), and then to a better customer retention (b), and hence to a better communication with customers (c). Even if the operational EB-objectives are basically the same, priorities are different (b-c-a for SEs and b-a-c for MEs).

After the introduction of EB in the own company, about the half of the LEs conducted a reorganisation of entrepreneurial processes (44,9%), while only 22,2% of MEs and 17,1% of SEs underwent such an important process to embed properly new technologies into the organisational structure and processes.

As far as Information Technology is concerned, ICT Budgets are somewhat in proportion to the entrepreneurial size: while 85,8% of SEs invest up to 25.000€ in the previous year, 81,9% of MEs invested up to 50.000€; LEs have a wider budget range, with a peak of 29% with values over 1 million € investment for EB-solutions. As far as EB-solutions are concerned, ASPs were chosen more frequently by SEs and MEs (respectively 20,1% and 18,5%), while more seldom by LEs (only 7,4%); as a matter of fact, 92,6% of LEs implemented EB-solutions, which are in 86,9% of the cases standard solutions with entrepreneurial adaptation or even complete individual solutions. In this respect, both SEs and MEs are opting preferably for simpler standard solutions without or with entrepreneurial customisation (79,3% and 68,9%). Furthermore, 93,3% of LEs and 81,1% of ME trained their employees to use specific EB-solutions, while only 65,9% of SEs did it; this fact means that 34,1% of SEs did not consider relevant to skill the personnel to make use of the adopted solutions.

3.3 Differences among economic sectors

Differences between different sectors of the economy (namely industry, services without commerce, and

commerce) were also sought with the help of bivariate and multivariate data analysis. Even if enterprises from the services sector do show to be slightly further within the EB life cycle, on the whole, the data do not underline any remarkable and significant differences in the pattern of use and deployment of EB among the considered sectors. As a result, it can be concluded that the importance and the degree of deployment of EB industry, services without commerce, and commerce can be considered as comparable.

3.4 Cluster analysis

On the base of a factor analysis, 15 factors were identified and hence used as variables for a hierarchical cluster analysis. The objective of the cluster analysis was to search for patterns to characterise different enterprise profiles.

Three clusters of enterprises have been identified: E-Business Experts, E-Business Followers, and E-Business Conservatives. The distribution of enterprises between the three clusters was, respectively, the following: 60 enterprises in the first, 165 in the second, 53 enterprises in the third cluster. The clusters are going to be discussed with the help of some selected data.

1. *E-Business Experts* (19,1% of all survey participants) are quickly moving towards the most mature phases of the EB life cycle. Among them, 24,5% have already deployed an embedded a holistic EB concept in the own organisation, while even 41,5% are using EB as a means for inter-organisational ICT-based integration. A considerable amount of enterprises (more than 50%) embedded EB thoroughly in their whole organisation – both in direct entrepreneurial functions (procurement, manufacturing, logistics, sales/ marketing, after-sales service) and in indirect entrepreneurial functions (research and development, personnel/ human resources) – and used EB also as an approach to integrate with their suppliers, partners and customers (SRM, SCM, and CRM).

Enterprises from this cluster have a remarkable EB-driven success: respectively 62,3% and 56,5% of the cluster enterprises could achieve either higher sales volumes or lower procurement costs through EB. Only 11,3, % and 17% did not have any positive effects in sales volume or procurement cost. Furthermore, at least 50% of the firms fulfilled or even exceeded their objectives. These results are on the whole clearly positive.

Members of this cluster have started with EB for four to five years, are innovative to highly innovative, and consider EB as important or even very important for the own business. For all above-mentioned reasons, the members of this cluster can be hence considered as the successful pioneers in the field of EB.

2. *E-Business Followers* (59,3% of all survey participants), the largest of the three clusters. In this group, a lot of the enterprises concentrates their activities on a web presence and web marketing (49,1%), while only 26,1% deployed or used basic EB-solutions, such as e-shops or e-marketplaces, and only a minority implemented a holistic EB concept or even approaches of inter-organisational integration (9,1%). Enterprises from this cluster have started with EB in average for three to four years and currently do concentrate their EB activates in about three entrepreneurial functions (mainly procurement, sales/ marketing, and after-sales service).

Even if these enterprises are not innovators in the field of EB, they tend to give EB a relative importance and seem to be aware of the potentials that can be reached with its embedding into their organisations. Even if less than the previous cluster, respectively 31,5% and 34,5% of the cluster enterprises could achieve either higher sales volumes or lower procurement costs through EB. Still, 30,9% and 32,7% did not reach any positive effects in sales volume or procurement cost.

In the future, members of this cluster plan to extend their EB activities: 37,6% plan to implement stand-alone EB solutions, while 29,1 plan to deploy a holistic EB concept in the own enterprise or more advanced inter-organisational integration; this is reflected by the fact that such enterprises plan to extend EB to entrepreneurial functions such as logistics, and CRM, but also personnel/ human resources. To conclude, for all above-mentioned reasons, members of this cluster can be also defined as second-movers.

3. *E-Business Conservatives* (21,6% of all survey participants), a smaller group than the previous one. Also in this cluster, the majority of enterprises (53,4%) concentrates their EB activities on web presence and web marketing, while 35% gained initial experiences with EB-solutions. Enterprises from this cluster have started with EB in average for less than 3 years and currently concentrate their EB activities only in one or two entrepreneurial functions (mainly procurement and sales/ marketing). Many of the members of this cluster can be considered as late followers, since they started to deploy EB only when the related solutions reached a considerable degree of maturity. It is striking that about 40% of the members of this cluster did not define any objectives of their own EB activities. This could be interpreted with the fact that many of such enterprises were pushed from the market (competitors and customers) to adopt initial EB solutions, and hence just reacted to the new boundary conditions without defining any further objectives. For the same reasons, enterprises of this cluster are also not particularly innovative, but follow the mature solutions offered in the market; furthermore, they do give EB only a neutral importance. As far as economic success is concerned, only 18,3% had a higher sales volume, and also only 18,3% could achieve lower procurement costs through EB. About

40% of the members of this cluster did not reach either higher sales volumes nor lower procurement costs at all. In the future, EB Conservatives plan to implement some EB solutions (25%) or to deploy and implement a holistic EB concept or even approaches of inter-organisational integration (26,6%); this fact is being reflected by the fact that they plan to extend EB to entrepreneurial functions such as logistics, after-sales service and CRM. All in all, members of this cluster can be also defined as cautious followers or third-movers.

3.5 Critical success factors for E-Business

Last but not least, several hypotheses were tested, in order to identify critical success factors for EB on the basis of the available data. The success factors that were identified are: 1) the existence of an EB strategy, 2) the existence of an organisational form to coordinate EB activities, 3) the execution of reorganisational measures, 4) the choice to implement individual or customised solutions, and 5) the decision to train employees. By the analysis of such factors within the three clusters, it could be observed the majority of the EB Experts were fulfilling all criteria.

7. CONCLUSIONS, LIMITATIONS AND FURTHER NEED FOR ACTION

In this paper the results of a survey with German enterprises in the field of E-Business were presented and discussed. Among others, three entrepreneurial profiles in the field of E-Business could be identified and characterised: successful pioneers, second movers, and cautious followers.

The survey was of an explorative kind and it is not necessary representative. Anyhow, interesting results and some trends could be highlighted. Hence, the survey can be seen as the initial step of a wider research framework. As a matter of fact, its ultimate objective was to enable an initial analysis of the research field, in order to enable the identification of patterns, trends and research areas to be then further and deeply analysed within following works of empirical research. The results support the fulfilment of this objective.

With the help of the gained data, a basis for further analysis was identified. The next step will be a set of semi-structured interviews to be conducted with targeted German enterprises from the first cluster (EB Experts) from selected industrial sectors. The objective is to better understand the reasons that lead to the described results, to deepen the analysis of critical success factors, and to highlight differences between industrial sectors.

ACKNOWLEDGEMENT

The research was conducted within the activities of the

“Aachen Competence Centre – Electronic Commerce (ACC-EC)”, a project (grant number: VI B4 –003068/4, see also www.aixplorer.de) funded by the German Ministry of Work and Labour.

ACC-EC is funded within the framework “Network Electronic Commerce” (“Netzwerk elektronischer Geschäftsverkehr”), which is a Germany-wide network of 33 local competence centres in E-Commerce and 7 federal competence networks that work in close contact with German enterprises in several fields related to E-Business (see also www.ec-net.de).

REFERENCES

- [1] ADM e.V. Richtlinie für Online-Befragungen, http://www.adm-ev.de/pdf/R08_D.pdf, Download on 02.06.2004.
- [2] Backhaus, K., Erichson, B., Wulff, P. Weiber, R.. *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung*, 10th ed., Berlin: Springer, 2003.
- [3] Bitkom e.V. *Daten zur Informationsgesellschaft - Status Quo und Perspektiven Deutschlands im internationalen Vergleich*, Berlin, 2004.
- [4] Bosnjak, M. (Non)Response bei Web-Befragungen, Dissertation of the Mannheim Univ., Aachen: Shaker, 2002.
- [5] Brosius, F. SPSS 8, 1998, (Download at 02.06.2004 from <http://psypcaum0.psych.nat.tu-bs.de/studium/manuale/spss/>)
- [6] Dannenberg, M.; Barthel, S. *Effiziente Marktforschung*, 1. ed., Bonn: Galileo, 2002.
- [7] Decker, D. *Marktforschung mit dem Internet, Einsatzmöglichkeiten, Grenzen und Entwicklungspotenziale*, Marburg: Tectum, 2001.
- [8] Hauptmanns, P.; Lander, B. Zur Problematik von Internet-Stichproben. In: Theobald, A.; Dreyer, M.; Starsetzki, T. (eds.) *Online-Marktforschung - Theoretische Grundlagen und praktische Erfahrungen*, 1st ed., Wiesbaden: Gabler, 2001, pp. 27-40.
- [9] Hüttner, M.; Schwarting, U. *Grundzüge der Marktforschung*, 7th ed., München: Oldenbourg Wirtschaftsverlag, 2002.
- [10] Kamenz, U. *Marktforschung - Einführung mit Fallbeispielen, Aufgaben und Lösungen*, Stuttgart: Schäffer-Poeschel, 1997.
- [11] Koch, J. *Marktforschung - Begriffe und Methoden*, 3rd ed., München: Oldenbourg Wissenschaftsverlag, 2001
- [12] Luczak, H.; Forzi, T. E-Business Perspectives, in: *Int. J. Internet and Enterprise Management*, Vol. 2., No. 2, pp. 2, 113-118
- [13] NFO Infratest GmbH & Co. KG. *Monitoring Informationswirtschaft, 6. Faktenbericht*, München, 2003. (Download on 02.06.2004 from <http://193.202.26.196/bmwi/index.htm>)
- [14] Unger, F. *Marktforschung - Grundlagen, Methoden und praktische Anwendungen*, 2nd ed, Heidelberg: Sauer, 1997.
- [15] Weis, H.C.; Steinmetz, P. *Marktforschung*, 5th ed., Ludwigshafen: Friedrich Kiehl, 2002.

Exploring the Reality of E-Commerce Benefits among South African Businesses

Alemayehu Molla

IDPM, The University of Manchester, Manchester M13 QH, UK

Alemayehu.molla@man.ac.uk

ABSTRACT

The use of e-commerce by businesses in developing countries is related to the potential benefits of improving linkage to international value chains, increasing market access and reach, enhancing internal and market efficiency and lowering transaction costs. The opportunity imperative has resulted in the adoption of e-commerce by some businesses in these countries. However, the questions of what and how much benefit businesses in developing countries are extracting from their e-commerce investments are not well covered. This paper attempts to explore the real benefits of e-commerce based on data from 92 businesses in South Africa. The findings indicate that e-commerce benefits are by and large limited to improving intra and inter-organizational communications. Strategic benefits such as improving relationships across the value chain, increasing market reach and reducing market, operation and supply chain management costs are not as widely dispersed as the standard model of e-commerce would have made us believe. These findings support the argument that cautions against an over-optimistic view of e-commerce for developing countries.

Keywords: E-commerce, E-commerce benefits, Developing countries, South Africa

1. INTRODUCTION

Many prognoses of e-commerce in developing countries emphasize the opportunities that e-commerce can unveil for firms in developing countries to easily access the global market, strengthen their linkage to global supply chains and the range of cost saving, disintermediation and competitive advantage benefits that would be expected to follow such access (Goldstein and O'Conner, 2000; UNCTAD, 2003).

Others question the feasibility (and general optimism) of e-commerce for developing countries based on untested extrapolations from the experiences of industrialized countries. For instance, Pare (2002) argues that the idea of efficiency gains promised by e-commerce is far-fetched and overlooks the lack of core capabilities. Based on case studies of Chinese organizations, Hempel and Kwong (2001) and Ho and Chen (1999) argue that there is a mismatch between the culture embedded in e-commerce. Odedra-Straub (2003) reinforces the above and opines that the opportunity imperative undermines the resource and local infrastructure realities of developing countries' businesses.

Therefore, understanding what and how much benefits businesses in developing countries are extracting from e-commerce investments is important. This paper aims to explore the real benefit of e-commerce for businesses in developing countries. First, we review the literature on the potential benefit of e-commerce in general and for developing countries in particular. Then, based on data extracted from a survey of South African businesses, we explore the benefits of e-commerce. We finalize the paper by discussing the findings and making some preliminary conclusions.

2. THEORETICAL BACKGROUND

Of the various potential benefits of e-commerce for developing countries argued in the literature, three broad themes can be identified. These are improving linkage and market access, disintermediation outcomes, and improving firm efficiencies.

2.1 Linkage and Market Access

In the era of increased internationalisation of goods, labour and information markets (Wigand et al, 1997), markets and more specifically access to markets remain one of the crucial problems of businesses in developing countries. Especially, with the dominance of global commodity chains, linkage to global supply networks influences not only access to markets but also firm competitiveness (Dolan and Humphrey, 2001; Moodley, 2003).

Information and communication networks have the potential to make location-related constraints less influential and provide a better worldwide access to markets and market information (Wigand et al, 1997). They also facilitate the integration of national trading systems to global systems (Gereffi, 2001). It, therefore, follows that using e-commerce, businesses in developing countries, irrespective of size and location, could overcome the geographical barrier of trading globally and could access markets and supply networks that would have otherwise been impossible to them. As a result existing firms would be able to increase their market 'reach' and new firms would be able to flourish.

2.2 Disintermediation Outcomes

Disintermediation refers to the death (or gradual elimination) of the middleman (the intermediary) from the market value chain as organizations rely on the

potential of electronic networks to establish direct linkages with consumers and suppliers (Wigand and Benjamin, 1995). Although the disintermediation effect of e-commerce on all types of intermediaries and in all types of industries is critically challenged (see for example Bakos, 1998; Sarkar et al, 1995), it is tenable to argue that networks could enable organizations to bypass (even if not to completely eliminate) some, if not all, of the intermediaries and reduce the producer-intermediary or the intermediary-consumer transaction costs.

On the other hand, businesses in developing countries including those in the agriculture sector often depend on long supply chains both to market their products and to purchase required inputs (Dolan and Humphrey, 2001). More often than not, the intermediaries take the lion's share of the profit and they decide which products are to be delivered to the market and from which supplier to purchase equipment and other necessary inputs. They also add to the cost of input materials and finished products. As indicated above, e-commerce can enable producers and/or consumers to bypass some of these intermediaries and/or the costs associated with them.

2.3 Efficiency Gains

One of the costs that significantly affect efficient performance of a business or any other economic activity is the coordination cost (Malone et al, 1989). Coordination costs include costs incurred in coordinating the activities of people and work and the costs of participating in the market (Wigand, 1995; Sarkar et al, 1995). Malone et al (1989) postulate that advances in information and communication networks improve the speed and cost of communicating the same unit of information and enable the design and deployment of strategic linkages among market players. It also facilitates an electronic market place where buyers and sellers easily compare offerings. This, in turn, leads to efficient utilization of resources in coordinating activities and in reduced costs of transaction (Malone et al, 1989). Wigand (1995) relates the above benefits of electronic networks to e-commerce and indicates that through e-commerce organizations would be able to achieve an even cheaper unit cost of coordinating activities.

It has been argued that businesses in developing countries incur high costs in both production and coordination of their economic activities because of inefficient systems of procurement, communication, inventory control and operation (Mann, 2001). Such high costs normally add to the market price of products and affect the competitiveness of most developing countries' products in the global market environment. Therefore, from the notion of the argument pursued above, it follows that through e-commerce, businesses in developing countries would be able to reduce the cost of coordinating the work of people and machines in all

the "information, negotiation and execution phases" (Wigand et al, 1997) of their systems and increase their efficiency. In addition, through e-commerce, firms in developing countries would be able to reduce the transaction costs they would have otherwise incur to participate in international trade and sell their products and services more competitively.

In the remaining part of the paper, we explore to what extent the businesses surveyed in South Africa realized the above-hypothesized benefits of e-commerce.

3. NOTES ON RESEARCH METHODS

The data used in this paper is extracted from a survey response of 150 businesses conducted as part of a doctoral research project. For the purposes of this paper though, the e-commerce capability of the businesses was used as a data extraction criterion. In order to talk about e-commerce benefits, we assumed that the businesses needed to have at least an informational e-commerce capability. As a result, only businesses that have attained an informational or interactive or transactional or integrated e-commerce capability were selected. This has resulted into 92 data subsets.

We drew a set of e-commerce benefit statements from the review of the literature. The respondents were asked to express their degree of agreement about each of the benefit statements on a five point Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree). We specifically sought out the CEOs, general managers and managing directors of the organizations to respond to the survey questions. The data were analysed using descriptive and exploratory statistical techniques.

A wide range of demographic characteristics was covered by the data subset. The majority of the respondents (72%) held a job title of managing director or CEO or general manager. The remaining respondents held directorial positions in finance (11%), IT (10%), e-commerce (4%) and marketing (3%). Eighty six percent of the organizations have been in business for more than 10 years. In terms of size (using the conventional measure of employee number), 83% can be characterized as large. Some 58% of the responses were from the manufacturing and services sectors. While the primary sector, i.e., agriculture, construction and mining represents 17% of the responses, trade and tourism and the ICT sectors account for 15% and 10% of the responses respectively.

4. FINDINGS

4.1 E-commerce Benefit Constructs

Exploratory factor analysis was used in order to identify underlying constructs and investigate relationships among the 16 items used to assess e-commerce benefits. The factors were extracted using an iterative sequence

of principal components extraction technique with eigenvalues greater than one. To facilitate interpretability, this was followed by varimax rotation. The criterion used for assigning an item to a factor was a minimum factor loading of 0.5. In addition, single item factors and items with a factor loading greater than 0.5 on two or more factors were excluded from subsequent iterations. The final factor structure with 3 factors containing 12 items was obtained after the third iteration (Table 1). In addition, Cronbach's alpha was counted and the scores were above the accepted level (Cronbach, 1979).

Examination of the results in Table 1 indicates that the first factor with six items accounts for 42.8% of the variance. We labelled this factor as the *market performance* gains of e-commerce and is related to both the linkage and market access and disintermediation benefits of e-commerce. Factor one appears to represent variables that constitute benefits in the downstream of the value chain. Businesses that use e-commerce can potentially increase their market reach and at the same time implement mass customisation strategies to produce products and services that suit the needs and preferences of individual consumers (Fruhling and Digman, 2000). Organizations can also use e-commerce to reengineer the selling and distribution processes and eliminate some of the intermediary activities to develop direct contact with their customers. This allows locking-in of customers, improving their relationship with the business and developing their loyalty. As a result, e-commerce might contribute to increasing the business bottom line and promoting competitiveness (Warrington et al, 2000). The results of our survey (Figure 1) indicate that market performance e-commerce benefits among the South African businesses significantly fall short of the above expectations.

The second factor with two items accounts for 14.7% of the variance and is labelled as *communications* improvement benefits of e-commerce. Kalakota and Whinston (1996) consider communication as one dimension of e-commerce. Use of e-mails, intranets and extranets might help organizations to improve the reach and richness of the information to be communicated. In addition, such networks could improve the speed and cost of communicating the same unit of information and enable the design and deployment of strategic linkages among market players (Malone et al, 1989). Our finding (Figure 1) indicates that this is one area that the businesses seem to reap the most out of e-commerce.

The third factor with four items accounts for 10.66% of the variance. This factor is related to the firm efficiency theme identified in the literature review and is labelled here as the *transaction cost reduction* benefit of e-commerce. Wigand and Benjamin (1995) differentiate between market and hierarchy transactions. While market transactions refer to those that support coordination between multiple buyers and sellers, hierarchy transactions, on the other hand, refer to those supporting coordination within the firm as well as the industry value chain. The items under factor three appear to show hierarchy transaction cost savings. Such savings are expected from reductions in operation costs (such as personnel, rent and order processing) supply acquisition and supplier management costs and improvement on internal processes. However, such benefits of e-commerce appear to have eluded our respondents (Figure 1).

Table 1: E-commerce benefits factor analysis

Variable	Factor Loadings (Varimax raw) (ECbenefit) Extraction: Principal components (Marked loadings are > .500000)		
	Factor 1	Factor 2	Factor 3
Extending firms reach (market)	0.840		
Product/service differentiation	0.846		
Increased customer loyalty and retention	0.809		
Improved revenue	0.596		
Improved competitive position	0.620		
Improved customer relationship	0.733		
Improved internal communication		0.859	
Improved external communication		0.906	
Improve process speed			0.588
Reduced operation cost			0.735
Reduced cost of purchasing and procurement			0.833
Improved supplier relationship			0.736
Eigenvalue	5.142	1.768	1.280
Total Variance explained %	42.85	14.73	10.66
Cronbach's Alpha	0.87	0.71	0.78

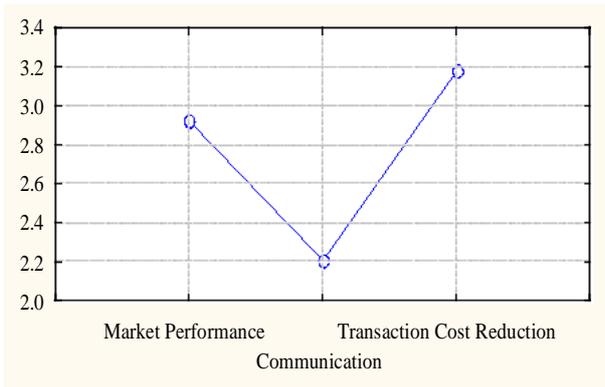


Figure 1: Status of E-commerce Benefits

The three factors extracted above demonstrate some relationship (Table 2). Particularly, there is a relatively stronger correlation between market performance and transaction cost reduction. This implies that if a business is not successful in reducing transaction costs, then it is unlikely for that business to perform better in the market, which means it would find itself at a competitive disadvantage.

Table 2: Correlation coefficients

	Correlations (Ecbenefit)		
	Market Performance	Communication	Transaction Cost Reduction
Market Performance	1.00	0.29	0.54
Communication	0.29	1.00	0.35
Transaction Cost Reduction	0.54	0.35	1.00

4.2 Impact of Demographic Variables

To explore differences of e-commerce benefits in small and medium and large enterprises, the mean score values of the three factors were plotted (Figure 2). Examination of Figure 2 indicates a more or less similar pattern. In both small and medium and large organization groups, improvements in internal and

external communications are reported by most organizations while cost reduction and market performance benefits remain the least experienced. The graph also indicates a relatively better e-commerce benefit for small and medium enterprises in the first two factors, i.e., market performance and communications. The significance of this difference is further tested using a one-way ANOVA. The result (*Wilks lambda*=0.967, $F(3, 83)=0.9573, p=.417$) indicates that the difference is not statistically significant.

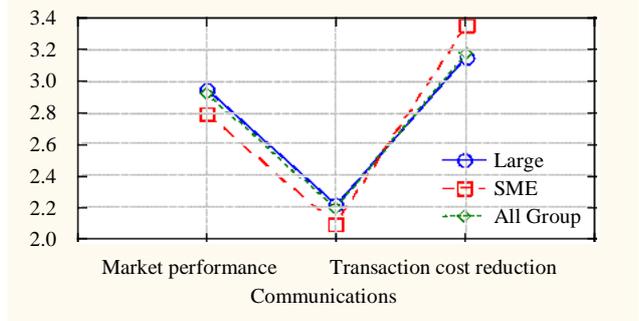


Figure 2: E-commerce benefits by business size

In addition to size, we tested if businesses belonging to different sectors demonstrate variations in terms of e-commerce benefits. The sector wise mean benefit scores are plotted in figure 3. The result has not produced a single pattern indicating one particular sector benefiting more from e-commerce than other sectors. The ANOVA test (*Wilks lambda*=0.88446, $F(12, 211.95)=0.83895, p=0.61022$) with no statistically significant result has reinforced this finding.

We anticipated the e-commerce capability of businesses to have some impact on e-commerce benefits. The mean plot (Figure 4) appears to indicate a clear distinction in terms of benefits between businesses with informational e-commerce capability and those that have developed integrated e-commerce capability. As anticipated, businesses that have developed an integrated e-commerce seem to have reported greater benefit from their e-commerce investment in all the areas. However, the ANOVA and discriminant function analysis tests did not produce statistically significant results showing the effect of e-commerce capability on the benefits of e-commerce.

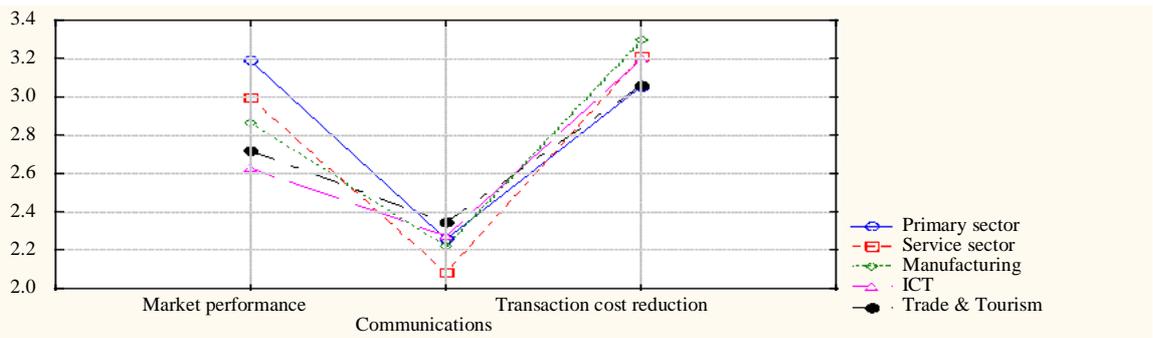


Figure 3: E-commerce benefits by sector

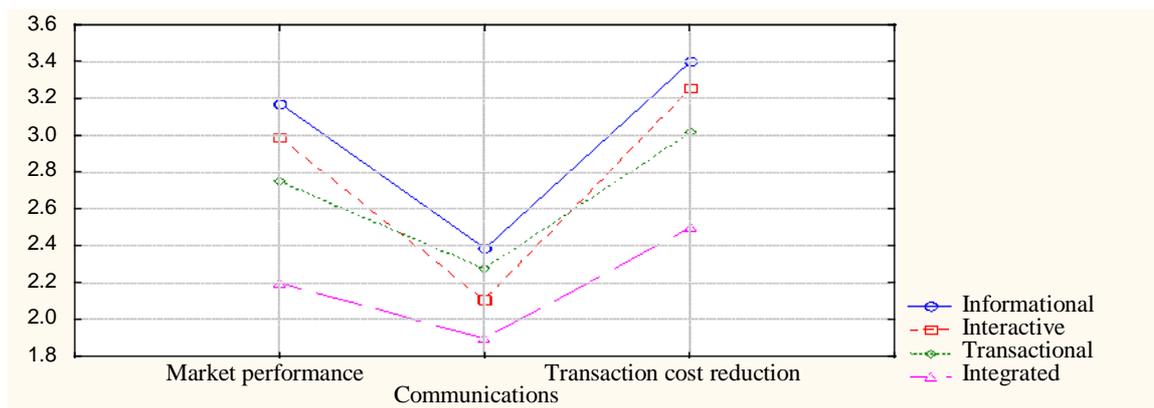


Figure 4: E-commerce benefit by e-commerce capability

5. DISCUSSIONS AND CONCLUSION

Our findings about the benefits of e-commerce for businesses in South Africa appear to contradict most of the themes outlined in the theoretical background section. They also appear to contradict to the buoyant optimism surrounding e-commerce's potential for developing countries. The majority of the businesses do not appear to have obtained e-commerce benefits in terms of expanding their market reach, improving their competitive position, and improving relationships with customers and suppliers. Cost reductions due to savings from improved internal processes and operations have not been as widely found as the standard model of e-commerce would have made us believe.

The results of the study reinforce many of the empirical findings of previous research on e-commerce in developing countries. For instance, Moodley et al (2003) based on a study of garment exporting firms in South Africa found that e-commerce has neither allowed those firms to by-pass middleman nor reduce trade coordination and transaction costs. Pare (2003) analysis of 117 B2B e-hubs discovered that neither are the e-hubs providing services that contribute towards transaction cost reduction nor are the businesses participating in those hubs reaping such benefits. Humphery et al's (2003) findings on the garment and horticulture industries of selected developing countries also suggest that many potential benefits relating to transaction facilitation and reach did not apply to the businesses they interviewed.

One variable that might have affected our results and perhaps the results of many other developing countries focused e-commerce studies is the embryonic nature of e-commerce there. Zhu and Kraemer (2002) have found a strong and significant relationship between the richness of e-commerce capability in terms of information, transaction, interaction and customisation and supplier connection and its value and impact on business performance. The majority of the businesses in our survey (68%) could be categorized as having an entry-level (informational and interactive) e-commerce

capability with only 26% and 6% achieving a transactional and integrated e-commerce status respectively. This is consistent with findings from Brazil (Tigre, 2003), China (Tan and Ouyang, 2004), Malaysia (Le and Koh, 2002) and South Africa (Moodley, 2003). Such capabilities would most likely support the conduct of basic communicational and informational but not transactional activities, hence limiting the extent of e-commerce benefits.

Further, the experience of developed countries suggests that at the early stage of e-commerce there was a significant gap between its anticipated and actual achievements (Marshall et al, 2000). The NNI (1999) supports this and suggests that expecting cost savings from e-commerce instantly is not realistic because benefits might take a long time before they actually materialize. In developing countries in particular, the marginal cost of adopting e-commerce tends to be much higher than in other environments. This is because adopting e-commerce requires additional investments to put in place basic automation and informatisation systems and business networks. Hence, in the short term, e-commerce might add to the transaction cost further delaying the realization of cost saving benefits.

Mansell (2001) and Patterson and Wilson (2000) argued that the capacity of DC businesses to achieve perceived benefits of e-commerce is dependent on their ability to reduce many facets of technological divides and improve the institutional arrangements that support the conduct of business. Likewise, elsewhere, we found that the e-readiness of developing countries' businesses to have significant impact on e-commerce success (Molla, 2004). In particular, we showed that the e-commerce governance model organisations put in place, senior management commitment, and human, business and technological resources all influence marketplace and cost saving benefits.

The findings documented above have some further implications. Many businesses in developing countries are encouraged to adopt e-commerce in order to take full advantages of this new business platform. The result

of this current study provides researchers and practitioners with initial guidance that can be used to manage expectations of returns on e-commerce investments. It also extends previous findings that caution against the over-optimistic view of e-commerce both in terms of organisation-specific gains and contributions to socio-economic development. Particularly, it indicates that entry-level e-commerce applications are unlikely to generate the potential benefits hypothesized in the literature. Rather, what is implied here is that, the capacity of developing country businesses to achieve the perceived benefits of e-commerce requires building up their internal organisational capabilities. It also requires improvements in the richness of e-commerce sites on all informational, transactional and interactive dimensions and the institutional arrangements that affect its conduct. The study also provides researchers with an initial set of factors and items that could be used in future research. This would facilitate an understanding of the antecedents and outcomes of e-commerce in developing countries.

To summarize, our result shows significant shortfall from the optimistic view of e-commerce. However, it is not tenable to extrapolate from this and make sweeping generalizations about either e-commerce in South Africa or the wider developing countries context. In fact, this current study is limited in terms of its sample coverage in both geographic and size terms. In addition, the results we observed here are likely to change in the future as businesses in developing countries accumulate experience and improve their organizational and e-commerce capability. Hence, in order to better understand the e-commerce phenomenon in developing countries, further research that expands both the geographic and sample size coverage of our research is recommended. Research is also recommended that not only assesses the benefits but also the antecedents that affect those benefits.

REFERENCES

- [1] Bakos, Y., "The emerging role of electronic market places on the Internet", *Communications of the ACM*, Vol. 41, No. 8, pp35-42, 1998.
- [2] Cronbach, L.J., "Test validation", IN Thorndike, R. L.(ed.), *Educational Measurement*, 2nd ed, Washington DC: American Council on Education, pp443-507, 1979.
- [3] Dolan, C., J. Humphrey, "Governance and Trade in Fresh Vegetables: The impact of UK supermarkets on the African horticultural Industry", *Journal of Development Studies*, Vol. 37, No. 2, pp147-76, 2001.
- [4] Fruhling, A.L., L.A. Digman, "The impact of electronic commerce on business-level strategies", *Journal of Electronic Commerce Research*, Vol. 11, pp1-11, 2000.
- [5] Gereffi, G., "Shifting governance structures in global commodity chains with special reference to the Internet", *American Behavioral Scientist*, Vol. 44, No.10, pp1616-1637, 2001.
- [6] Goldsten, A., D. O'connor, *E-commerce for Development*, Retrieved October 5, 2000 from www.oecd.org/dev, 2000.
- [7] Hempel, P.S., Y.K. Kwong, "B2B e-Commerce in emerging economies: i-metal.com's non-ferrous metals exchange in China", *Journal of Strategic Information Systems*, Vol.10, pp335-355, 2001.
- [8] Ho, C., J. Chen, "The prospects of B2B e-commerce in China's Internet industry", *Business Forum*, Vol. 24 No.3/4, pp62-69, 1999.
- [9] Humphrey, J., R. Mansell, D. Pare, H. Schmitz, *The Reality of E-commerce with Developing Countries*, Media @LSE, 2003.
- [10] Kalakota, R., A.B. Whinston, *Electronic Commerce: A Manager's Guide*. London: Addison Wesley Publishing, 1996.
- [11] Le, T.T., A.C. Koh, "A managerial perspective on electronic commerce development in Malaysia", *Electronic Commerce Research*, Vol. 2, No.1-2, pp7-29, 2002.
- [12] Malone, T., J. Yates, R. Benjamin, "The logic of electronic markets", *Harvard Business Review*, May-June, pp 166-172, 1989.
- [13] Mann, C.L., *Global E-commerce: Issues and Policy for Developing Countries*. New York: Institute for International Economics, 2001.
- [14] Mansell, R., "Digital opportunities and the missing link for developing countries" *Oxford Review of Economic Policy*, Vol. 17, pp282-295, 2000.
- [15] Marshall, P., R. Sor, J. McKay, "An industry case study of the impacts of electronic commerce on car dealership in western Australia", *Journal of Electronic Commerce Research*, Vol. 1, No.1, pp1-16, 2000.
- [16] Molla, A., "The impact of eReadiness on e-commerce success in Developing Countries: Firm level evidence", Development Informatics Working Paper 18, IDPM: The University of Manchester, Available from <http://idpm.man.ac.uk/publications/wp/di/index.shtml>, 2004.
- [17] Moodley, S., M. Morris, M. Velia, "E-commerce for exporting garments from South Africa: 'digital dividend' or leap of faith?", *IDS Working Paper 182*, Institute of Development Studies, 2003.
- [18] Moodley, S., "Whither business-to-business electronic commerce in developing economies? The case of the South African manufacturing sector", *Information Technology for Development*, Vol.10, No.1, pp25-40, 2003.
- [19] NNI-Nolan Norton Institute, "Electronic commerce: The future is here", *Research Report*, Retrieved September 2, 2000 from www.kpmg.com, 1999.

- [20] Odedra-Straub, M., "E-commerce and development: Whose Development?", *Electronic Journal of Information Systems in Developing Countries*, Vol.11, No.2, pp1-5, 2003.
- [21] Pare, D., "B2B e-Commerce services and developing countries: Disentangling Myth from Reality", *Association of Internet Researchers International Conference*, October 13-16, Netherlands, 2002.
- [22] Pare, D., "Does this site deliver? B2B e-commerce services for developing countries", *The Information Society*, Vol.19, No.2, pp123-135, 2003.
- [23] Patterson, R., E.J. Wilson, "New IT and social inequality: Resetting the research and policy agenda", *The Information Society*, Vol. 16, No. 1, pp77-86, 2000.
- [24] Sarkar, M.B., B. Butler, C. Steinfield, "Intermediaries and cybermediaries: A continuing role for mediating players in the electronic marketplace", *Journal of Computer Mediated Communication*, Vol.1, No.3, pp1-13, 1995.
- [25] Tan, Z., W. Ouyang, "Diffusion and impacts of the Internet and e-commerce in China", *Electronic Markets*, Vol. 14, No.1, pp25-36, 2004.
- [26] Tigre, P. B. "Brazil in the age of electronic commerce", *The Information Society*, Vol.19, No.2, pp33-43, 2003.
- [27] UNCTAD, *E-commerce and Development*. Retrieved September 2002 from www.unctad.org/e-commerce, 2003.
- [28] Warrington, T.B., N.J. Abgrab, H.M.Caldwell, "Building trust to develop competitive advantages in e-business relationships", *Competitiveness Review*, Vol.10, No.2, pp160-168, 2000.
- [29] Wigand, R.T., R.I. Benjamin, "Electronic commerce: Effects on electronic markets", *Journal of Computer Mediated Communication*, Vol.1, No.3, 1995.
- [30] Wigand, R., A. Picot, R. Reichwald, *Information, Organization and Management: Expanding Markets and Corporate Boundaries*, Chichester, England: John Wiley & Sons, 1997.
- [31] Wigand R.T., "Electronic commerce and reduced transaction costs: Firms' migration into highly interconnected electronic markets", *Electronic Markets*, Vol. 16/17, pp1-5, 1995.
- [32] Zhu, K., K.L. Kraemer, "E-commerce metrics for net-enhanced organisations: Assessing the value of e-commerce to firm performance in the manufacturing sector", *Information Systems Research*, Vol.13, No.3, pp275-296, 2002.

Intended Usage of Online Supermarkets: The Singapore Case

Hui Tak Kee, David Wan

NUS Business School, National University of Singapore, Singapore 117592
{bizhuitk, bizwantw}@nus.edu.sg

ABSTRACT

The online grocery sales experienced a surge in Singapore especially in the early half of 2003 during the SARS outbreak. The Technology Acceptance Model and some basic demographic variables will be used to study the intended usage of online supermarkets. Based on a sample of size 211 shoppers, it was found that those 21-40 years old have a propensity in using the online grocery. However, those having income lower than \$1,500 have less inclination in using the e-grocery. The stepwise discriminant analysis shows that two perceived variables, usefulness and ease of use, and two demographic variables, age, income are significant in differentiating the intended customers. Our results are different from the past studies where the gender and education are found to be significant. This may be due to the fact that most respondents in the sample are under 41 years old with at least tertiary and above education.

Keywords: Online supermarket, factor analysis, discriminant analysis

1. INTRODUCTION

A supermarket is a large self-service food store (McClelland, 1962). As the grocery industry grew, this definition also expands to encapsulate items such as basic household requisites although its strongest association remains with groceries. Priluck (2001) revealed that consumers in US make an average of 2.2 visits to a supermarket per week to shop for groceries which is signaling the importance of such an activity.

However, grocery shopping has taken on a new dimension with the advent of the Internet. Riding on the rising trend of online retailing, supermarkets and logistics companies have equally come out with their own version of online supermarkets. Lorek (2001) indicated that online grocery sales seven years ago in the United States were only \$85 million but it was estimated that the online grocery store revenue may reach \$1.3 billion in 2002. Today, traditional brick-and-mortar stores are increasingly entering the online market in an attempt to leverage both brand awareness and create distribution (Kempiak et al., 2002).

Nonetheless, not all is rosy for the players trying to create their own sphere in the e-grocery businesses. Britain's No. 1 supermarket chain – Tesco – has watched one rival after another put up the white flag on their e-grocery businesses (The Business Week, 2000). In the U.S., the failure of the Webvan Group Inc. in trying to establish a purely Web-based grocer was fuelled by the low customer demand that was not sufficient for Webvan to operate the facilities at anywhere near their capacity.

Given the mixed success signal sent by online grocery retailers, it is vital for supermarket operators to understand how people decide whether they will use the grocery services provided online, for systems that are not used cannot be effective no matter what their

technical merits. This need is particularly relevant for the increasingly competitive online grocery retail market, in which numerous supermarkets compete among themselves within a relatively stable market (Rohm et al., 2004). Furthermore, from a managerial perspective, classification of potential users provides the basis for understanding and targeting different groups of consumers.

Given that online grocery shopping is nascent in Singapore, a typology specific to this medium of purchasing will enable us to identify distinct segments of potential consumers, thereby allowing retailers to effectively tailor their offerings to these customer types. Also, such a new retail channel cannot be effective unless they are used. Consequently, it is important for supermarket operators to understand their consumers' attitudes toward using online services. Our objective is thus to identify and profile the characteristics of such potential users of online supermarket and in turn, make a preliminary assessment of consumer demand for online grocery shopping services based on attitude measures.

More importantly, we have attempted to use a relatively new model, the Technology Acceptance Model (TAM), which serves to articulate the core psychological aspects associated with technology use. Based on a generic model of attitude and behavior, the TAM has provided a robust and valuable model when considering information technology acceptance by potential mainstream users (Taylor and Todd, 1995). It has also been able to provide good prediction of an individuals' intention to use a particular information system.

To date, Singapore enjoys reliable and sophisticated networks for information technology and telecommunications services, transportation and utilities. The information technology and

telecommunication infrastructure of the city-state is particularly well developed. During the height of dot-com boom, 47% of households have personal computers while 53.8% of the residential populations have Internet access or subscriptions. In 2003, online grocery sales experienced a surge, particularly in the early half of the year during the SARS outbreak. This was apparently good news for NTUC Fairprice and Cold Storage, the two main supermarkets in Singapore that offer online grocery shopping. It was reported that online sales of Fairprice Online have attained an increase of as much as 300% in one week at the peak of the outbreak.

It is our goal to construct a forecasting model to predict the intention of consumers to shop for groceries online, which in turn could give a reasonable estimation to actual system use. Specifically, we want to look at consumers who have access to the Internet, and to seek an understanding of their attitudes toward online grocery services. In short, we would like the prediction model that we have established to forecast an individual's intention to use the online service.

The paper is structured as follows. First, we take a look at the present situation of supermarkets in Singapore in the introduction section. Second, we review the literature on online grocery shopping and the TAM. Third, we discuss the methodology and data collection procedure. Fourth, we present our analyses and the results. Lastly, we discuss the limitations and the implications of this study, and draw a conclusion.

2. LITERATURE REVIEW

2.1 Profile of E-grocery Shoppers

Online grocery shopping accounted for less than 1% (\$150 million) of the total supermarket sales (\$440 billion) in 1998 (Beck, 1999). Sales of online grocers concentrated around specialty foods and it is likely that this trend will continue. In fact, market analysts expected that 5% to 8% of all groceries sales would transact through online within the next five years (Lundegaard, 1997; Kirsner, 1999). Furthermore, Anderson Consulting (Buss, 1999) had predicted the online supermarket market in US to increase significantly from 200,000 people currently to 20 million in 2007. With the increasing prevalence of online shopping in the long run, it is easy to understand the industry's interest in online grocery shopping. A recent forecast made in a Forrester research report predicted that online retail sales would increase from US\$45 billion in 2000 to US\$269 billion in 2005, which is a growth of close to 600 percent (Dykema, 2000). The study by Buss (1999) noted that even by achieving 1% of the grocery market would translate to a US\$1 billion per year business for online grocers.

There is an extensive range of literature that can be found with relation to online grocery shopping. In particular, various demographical studies had been carried out to identify the various groups associated with online grocery shopping. It has been pointed out that such online shoppers tended to be "boomers" with large households, with their demographic profile skewed towards the 35-49 age groups. In addition, those who shopped for groceries online were generally well educated, with three-quarters having obtained at least a college degree. In a study conducted by Park et al., (1998), they categorized online grocery shoppers into two groups: hi-tech baby boomers and older/ low-income suburban families. Not surprisingly, a survey carried out by Priluck (2001) also revealed that education was the main demographic variable affecting the tendency to buy groceries online, whilst the other demographic variables proved to be less significantly related to consumers' willingness to purchase grocery items online.

A similar study by Hiser, Nayga and Capps (1999) also found that variables such as income, the number of people living in the household, the presence of children and gender were not significant determinants to the usage of the online grocery shopping service. Age and education have proved once again to have a positive relationship in this study. However, it also revealed that people over age 50 were less likely to consider the service, as compared to people from age 18 to 29 years old.

These studies on online grocery shoppers' demographics provide the basis of understanding and targeting different customer groups. Furthermore, through the identification of definite online shopper segments, retailers can strategize their services to tailor to the needs of specific segments.

2.2 Technology Acceptance Model

In an attempt to better understand user acceptance of new information systems, Davies et.al. (1989) developed the Technology Acceptance Model (TAM). They postulated how users' attitudes toward using a computer system influences the actual use, or acceptance of this system. The two major design features are the 'perceived usefulness' of the system, and the 'perceived ease of use'. The former is defined as the "degree to which an individual believes that using a particular system would enhance his or her job performance. The latter is defined as the "degree to which an individual believes that using a particular system would be free of physical and mental effort" (Davies, 1993). It was argued that these two features formed the users' attitudes toward using a computer system.

Research in social psychology has shown that behavior is best predicted by an individual's attitude toward the

behavior. For instance, it is difficult to estimate eventual system use if the system does not exist yet. Still, it was widely believed that an individual's intention to use the system could be measured. There is also considerable evidence that intention to perform a behavior predicts actual behavior (Sheppard et al., 1988).

Davies (1989) found that the TAM was able to predict the intention to use a word processing package fairly well. Henderson and Divett (2003) have also successfully applied the TAM to an online supermarket setting, thus providing empirical support for the ability of the TAM to predict actual behavior. It is the latter case that propelled us to adopt this model in predicting potential consumers' intention to use the online grocery services.

3. Methodology

3.1 Data Collection

The research sample employed in this empirical study consists of customers of two supermarkets in Singapore, namely NTUC-FairPrice and Cold Storage. Data were collected on February 2004 at each respective supermarket. A sample of 211 shoppers was taken altogether. These are customers who have just made their purchases from the supermarkets. Furthermore, we administered our questionnaires to those shoppers who have Internet access at home. The purpose of targeting only at grocery shoppers who have Internet access at home allows us to gain a specific understanding of the propensity of these actual grocery shoppers to shop for groceries online. Hence this group of grocery shoppers forms our target population. Furthermore, we only approach every third person that come out of the supermarkets to ensure randomness.

The majority of the respondents are below 40 years old (81.5 percent). Other than that, the profiles of the respondents are generally equitable, with approximately 50 percent of them being male and possessing some form of tertiary education. A total of 54 percent reports an annual income of \$2000 or below.

3.2 Measures

The survey is divided into three sections. The first section requires the respondents to provide a broad detail of their socio-economic backgrounds. In section two, we present two dichotomous questions which are used to elicit their responses toward their use of and intention to use online supermarkets. As the e-grocery services in Singapore just take off, we postulate that the number of online grocery shoppers has yet to reach a critical mass. This is bore out by the fact that only two of our respondents have ever used the Internet to purchase groceries. It is thus appropriate for us to get acquainted with the consumers' intention to use, as

opposed to seeking consumers who have used such services. This is supported by the TAM model as explicated previously.

Finally, a set of multiple-item scales is constructed under two main variables which are guided by the TAM framework. The questionnaire contains 6 attitudinal statements on a five-point Likert scale anchored by strongly agree to strongly disagree. These items examine the consumers' attitudes and perceptions toward the use of online grocery services

Contingency analysis is first used to test whether the demographic variables are associated with the intention variable. This will at least give some ideas on which demographic characteristics are more inclined to the use of the internet for grocery.

Factor analysis is then used to classify the initial six variables derived from the attitudinal measures in the questionnaire into two distinct factors, namely 'perceived usefulness' and 'perceived ease of use' using principal component and varimax rotation. The resulting factor scores will be used for the subsequent analyses.

Using the factor scores from the two derived factors and the profile of the respondents, we proceed to use the discriminant analysis technique to evaluate whether the consumers' perceptions of Internet as a shopping medium for grocery purchase and their profile characteristics can be used to predict their intention of using e-grocery services in the next one year.

In this study, we have identified 6 predictor variables (based on the demographical profiles and the TAM model) to predict the intention of an individual to use the e-grocery services. Furthermore, we use only 190 responses as an estimation sample to construct the classification rule to enable us to perform a post-hoc test of our model later. Lastly, we would classify the remaining 21 hold-out cases according to our discriminant model in order to assess the predictive validity of the model, that is, to test whether our model that is estimated with one set of data continues to hold good for the hold-out data.

4. RESULTS

Based on the four demographical profiles against the intention to use variable, the chi-square tests derived from the contingency analysis only shows the presence of a relationship between the three profile variables, namely, age, education level, income level, and the intention to shop for groceries online (all p-value < 0.05), and not the gender variable (p-value = 0.710 > 0.05). We can thus conclude that the intention to use this service is dependent on the age, education and income levels of consumers.

From the same results, we also notice that generally, those around 21-40 years old have a higher propensity to use online grocery services. This is expected since this group is the most economically stable. On the same note, those who have some form of tertiary education also express their intention to use such services in the near future. Those who possess income of less than \$1500 tend to have little inclination to using e-grocery services. As mentioned previously, the chances of consumers of any gender having the intention to use such services are almost equal. This initial profiling of the variable relationship would provide a basis for the affirmation of the results in the later analyses.

Two factors were extracted using Principal Component Analysis and rotated using Varimax with Kaiser Normalization. About 74.8% of the result can be explained by the two components, namely, the 'perceived usefulness' and 'perceived ease of use'.

The factor scores derived from the above factor analysis for the components of *Perceived usefulness* and *Perceived ease of use* are included as independent variables in the stepwise discriminant analysis. Other independent variables consist of Gender, Age, Income level and Education level. The Wilks' Lambda (p-value = 0.00) shows that the discriminant function is highly significant. The success rate for the predictions of membership of the criterion grouping variable's (i.e. intention to shop for groceries online) categories using the discriminant function developed in the analysis is 87.4%. The classification results also reveal that subjects with no intention to shop for groceries online are accurately classified for most of the cases.

5. DISCUSSION

Our prediction model obtained through discriminant analysis is able to predict the intended usage of online grocers with an accuracy of 87.4%, given their demographic and psychographic profiles. It is interesting to note that the discriminant analysis has yielded a different result from those of previous studies in our literature review. From the results, it is observed that the variables Gender and Education level are removed from the analysis. This result is deviated from those put forth in past literature (Hiser et.al. 1999; Priluck, 2001) which concludes that Age and Education level are the main variables in the determination of online grocery shoppers. These studies find that online grocery shoppers tend to skew towards a certain age group and that they are mostly well-educated, having obtained at least a degree.

The inclusion of Income level and the exclusion of Education level from the discriminant analysis can be explained from observations noted from the survey data from a sociological perspective of our local context. An analysis of the survey data reveals that

most of our respondents under 41 years old with tertiary or above education level and, also those with income level below \$1500 do not intend to purchase their groceries online in the next one year. These respondents are made up mainly of tertiary educated students who have no necessity to purchase groceries on a regular basis. Their parents would purchase the groceries for the family in most cases. Even if this is not so, this group of respondents is not constrained by the lack of time and family commitments. Also, it is observed that a significant number of working professionals under the same age group with tertiary education level and above, have intentions to use the online grocery shopping service in the next one year. In-depth analysis of their survey responses reveals that their main motive for intending to shop for groceries online is the convenience and timesaving factors online grocers provide. The main differentiating factor between the above two groups of respondents is their income level, as the later group of respondents earns a monthly income of \$1500 and above. Therefore, this results in income level having a greater influence than education level in the determination of intention to shop for groceries online.

Our study has some potential limitations. First, given the few variables identified, the respondents might be given too few options to indicate their preferences over the choices provided. It is likely that there exist other important variables that could explain one's propensity to shop for groceries online or predict one's intention. These variables, such as price or Internet accessibility, have not been studied in this paper and thus constitute a limitation of this research. More variables should have been included to produce a more comprehensive and robust model in order to capture the variance that is not explained by the variables we have identified.

6. CONCLUSION

There is considerable evidence that intention to perform a behavior predicts actual behavior (Sheppard, Hartwick, and Warshaw 1988). Under this premise, the model that we have constructed using discriminant analysis validates the empirical support for the prediction of intention towards the use or acceptance of electronic supermarket. Other than attitudinal influence, the results have also illustrated how certain demographic variables such as age and income level significantly influence users' perceptions and intention towards the usage of online supermarket. These variables have helped us build a good forecasting model that provides good prediction of individual's intention to use the online grocery system. Our findings also suggest that consumers will be motivated to shop for groceries online if they perceive the services offered by the online grocer to be useful and easy to use. This has implications to retail and marketing managers who are intending or have already set up an online grocery business.

From the results reported, marketers can target the specific age and income groups that show a high tendency to shop for grocery online. They could design their online grocery service to cater to the needs of these consumers, such as providing a large variety of products or making the online shopping process quick and hassle-free. Furthermore, they could utilize the reported results to interpret the factors that are keeping other consumers from using such a service, e.g. online grocery shopping is not convenient and time-saving, and examine ways to improve their service to attract a greater proportion of other consumer segments.

Some of the limitations of this research would provide a premise for future research. Although our research has identified a few main variables influencing the motivations of consumers to shop for groceries online, however there is also a need to address some other key factors, such as the quality of product and price that are associated with home shopping, which will also have an impact on a consumer's perception.

Online grocery shopping is a relatively new phenomenon in Singapore. The characteristics of the sample used in our research may change when the trend starts to be widely adopted, signaling the need for future research. In addition, most of the past studies are conducted and analyzed in the U.S. context, which might culturally differ from the Singaporean context. Therefore, it is recommended that future research explore the factors associated with the possible mediation effect of usefulness upon ease of use in our local context.

REFERENCES

- [1] Buss, D.D., *Damaged goods: Online Grocers Still Struggle to Deliver*, Business 2.0, Vol. 74, No. 8, 1999.
- [2] Davis, F.D., R.P Bagozzi and P.R. Warshaw, *User Acceptance of Computer Technology: A comparison of Two Theoretical Models*, Management Science, Vol. 35, 982-1003, 1989.
- [3] Davis, F.D., *User Acceptance of Information Technology: System Characteristics, User Perceptions and Behavioral Impacts*, International Journal of Man-Machine Studies, 38, 475-487, 1993.
- [4] Dykema E.B., *Online retail's ripple effect*, Forrester Report, 2000.
- [5] Henderson R. and M. J. Divett, *Perceived usefulness, ease of use and electronic supermarket use*, International Journal of Human-Computer Studies, Volume 59, Issue 3, pp 383-395, 2003.
- [6] Hiser, J., R.M. Nayga and O.Jr Capps, *An exploratory analysis of familiarity and willingness to use online food shopping services in a local area of Texas*, Journal of Food Distribution Research, Vol.30, No.1, pp.78-90, 1999.
- [7] Kirsner S., *Express lane*, Wired, 7, 112-112, 116, 118-120, 122, 1999.
- [8] Kempiak M. and M. A. Fox, *Online Grocery Shopping: Consumer Motives, Concerns, and Business Models*, First Monday, Vol. 7, No. 9, 2002.
- [9] Lorek L., *Online Grocery Business Still Ripe*, Interactive Week (16 July), pp. 34, 2001.
- [10] Lundegaard, K. M., *New owner to streamline Shopping Alternatives*, Washington Business Journal, 1997.
- [11] McClelland W.G., *Economics of the Supermarket*, The Economic Journal, Vol. 72, No. 285, pp. 154-170, 1962.
- [12] Park K., D. Perosio, G.A. German and E. W. McLaughlin, *What's in Store for Home Shopping?*, Cornell University Food Industry Management Program, Ithaca, NY, 1998.
- [13] Priluck R., *The impact of Priceline.com on the grocery industry*, International Journal of Retail & Distribution Management, Vol. 29, No. 3, pp. 127-134, 2001.
- [14] Rohm A.J. and V. Swaminathan, *A typology of online shoppers based on shopping motivations*, Journal of Business Research, Vol. 7, No. 7, pp. 747-757, 2004.
- [15] Sheppard B. H., J. Hartwick and P.R Warshaw, *The Theory of Reasoned Action; A Meta-Analysis of Past Research with Recommendations for Modifications and Future Research*, Journal of Consumer Research, Vol. 15, pp.325-343, 1988.
- [16] Taylor S. and P.A. Todd, *Understanding information technology usage: a test of competing models*, Information Systems Research, Vol. 6, pp. 144-176, 1995.
- [17] The Business Week Online [<http://www.businessweek.com>]

Positioning of Firms on an E-Business Value Chain: the Case of Video on Demand

Nabyla DAIDJ, Olivier EPINETTE, Pierre VIALLE

STRAT&MARKETIC Research Group, Institut National des Télécommunications, 91011 EVRY, France
{nabyla.daidj, olivier.epinette, pierre.vialle}@int-evry.fr

ABSTRACT

The provision of e-business services is supported by the implementation of value chains based on co-operation between firms combining different resources and capabilities in order to design, produce and market new products and services. The emergence of such value chains is characterised by the positioning of various players according to their specific resources and capabilities, whether they be existing or new firms. Our interest, by combining insights from the Resource-Based View and other perspectives, is to analyse according to what resources are necessary for actors to be able to successfully position themselves on a value chain, with a particular focus on relational assets, in the specific Video On Demand sector (VOD). This case is particularly relevant to show the structuring of a new value chain around new and “traditional” stakeholders. We shall first define the specific features of the movie industry. Secondly, we shall then present on the VOD value chain. Then we analyse how a company (called Moviesystem) has positioned itself on VOD value chain thanks to its specific resources and capabilities. We shall conclude by drawing general lessons from this case.

Keywords: e-business, Video on Demand, value chain, strategy, resources, relational assets

1. INTRODUCTION

The emergence of the so-called multimedia industry, based on the expected merging of three vertical industries (telecommunications, computing and the media) into a new horizontally layered one, raised tremendous interest at the end of the 1990s. There was a great deal of conjecture about how this new industry would be structured, and what types of player would be successful in positioning themselves in this emerging value chain ([6], [15]). A decade later, it seems interesting to analyse how value chains are finally structured and how and why some players have been able to position themselves successfully. Such a long delay is justified for two main reasons. First, the capabilities of the technical objects and systems at the time were clearly over-estimated, as was their speed of diffusion. In fact, it is only recently that adequate techniques and resources have become available to users. Second, different organisational arrangements and business models have been tested by various types of players, and time is therefore necessary in order to be able to draw lessons from these experiences.

This is particularly the case Video on Demand (VOD). On-line movies were probably the best-publicised example of multimedia service, and high business expectations gave rise to significant strategic moves from players of the three converging industries. VOD may be defined as content offered when demanded by consumers. Any video may be downloaded on a Personal Computer via the Internet. This service can be described very simply by the following value chain: it is developed by movie studios, bought by companies which package them and transform them into IP content, which are then sold to consumers either directly over a web site or

indirectly, through ISPs acting as distributors ([5], [10]).

Paradoxically, this business still remains largely under-developed among other on-line services. According to Jupiter Research, the world market for Video on Demand (VOD) amounted to only \$ 349 million in 2003, mostly in the United States. However, this market now appears promising. Until recently the bulk of the market was concentrated in the United States and was mainly served by cable operators. With the widespread uptake of ADSL connections, this market is expected to grow, and particularly outside the United States. The main ISPs are currently launching or preparing on line TV services that are or will soon be coupled with VOD services. According to Jupiter Research, The world market should reach a level of \$ 2 billion in 2006.

From our perspective, this case is especially relevant because it is a genuine case of convergence between three industries with different cultures, visions, business models and resources. Of particular interest is how a very specific industry has had to adapt to the constraints of electronic distribution. The provision of VOD services involve some significant adaptations of the movie value chain, such as content digitalisation and formatting, digital rights management (DRM), server platform management, movie catalogue management, relations with network operators or ISPs, or billing and customer management ([5], [10]).

We shall present how a VOD company, Moviesystem has positioned itself on the VOD value chain because of its specific resources and capabilities. These resources and capabilities allow Moviesystem to link the movie environment with the electronic distribution environment.

It is a good example of a company pioneering new forms of movie distribution by working with established film studios. Distributors are faced with strategic issues in terms of promotion and the associated time and geographic span of exhibition. Electronic distribution is a new step but it is also affected by the profit windows for film rights and by the distribution of the ownership of the film rights.

In this paper, we first of all present the specific features of the movies industry. Secondly, we will introduce the VOD market, the technological systems, the value chain and the challenges in providing VOD services. Then, we analyse how Moviesystem has positioned itself on the VOD value chain, in order to outline the critical resources that allowed Moviesystem to reach its current position. We conclude by drawing more general lessons from this case.

2. VIDEO ON DEMAND

2.1 The economical and technological context

VOD is a service providing video programs (typically movies, but other programs as well as requested by the customer in a catalogue and delivered at a time chosen by the customer. VOD is an expensive service, its cost depends on: the size of the program catalogue, the flexibility offered as to viewing and the quality of pictures sent. The VOD market must be put in perspective with the economics of the movie and media industry.

Movies are distributed through multiple channels in sequential release windows in a specific order. The majority of the revenues generated by a film are received within 18 months of a film's distribution cycle. After exhibition in its home country, the film passes over the following months into other channels: exhibition abroad, home video (cassette or DVD), Pay-Per-View, pay TV then free broadcast TV, leading to the profit release windows. This timeline must be respected by all stakeholders of the sector. This film cycle is based on its potential revenue along the different market segments, according to territoriality (by country and by linguistic zone) and time (duration and distribution rights) agreements [7].

VOD appears as a new release window in this film's distribution cycle. The question is to know when motion pictures will be available for VOD: should VOD share release dates with PAY-TV? Many VOD actors (in particular aggregators) insist on the necessity to link the VOD time window with the traditional and physical channel of delivering movies that are video rental/sell stores¹. It should be a condition of VOD success [11].

¹ "Some 80% of a title's home-video rental and DVD purchase business is done within the first 30 days of its hitting the rental shelves. In other words, the newer a movie, the more valuable it is. The implication is that the closer the VOD release window gets to the video rental window, the more VOD operators can

As broadband offers new opportunities for the distribution of movies, the development of VOD interests several actors: telecommunications operators (opportunity for them to adopt a «triple play»² strategy), ISPs, ADSL providers, television channels, cable and satellite operators, film companies and production studios. Operators see VOD applications as a means to increase the return on their broadband investments. TV channels can take advantage of telecommunications deregulation in order to sign partnerships with alternative service providers. Movie studios and film companies consider VOD as the missing link for an accelerated and a more efficient marketing and distribution of the films. They try to attract service providers through their huge catalogue of movies and programming.

Today, end users have mainly three options for viewing VOD programming [4]:

- Pay-Per-View / Near Video On Demand on cable/satellite systems: Video technique used by broadcasters using broadband distribution mechanisms such as cable and satellite TV. This form of broadcasting uses large amounts of bandwidth and is only viable for operators with large amounts of spare capacity. Cable Multiple system operators (MSOs) are particularly interested in VOD looking for a return on their investments in upgrading their networks to support digital services. These services are well positioned in particular in the countries where cable is fully deployed (USA)
- Personal video recorder (PVR) on cable/satellite systems: in the USA, Tivo and Sony Digital recorder, and in France Pilotime/Canal Satellite. PVRs are the first step in introducing a home media server. These PVRs (with several functionalities such as stop, pause etc.) allow customers to download a movie during the day in order to watch it a few hours later (in the evening). This offer relies on a more limited range of movies.
- Peer-to-Peer systems characterised by low quality, very large catalogue and variable viewing flexibility.

2.2 The VOD value chain

The VOD value chain will depend on the nature of technical architecture used and on the complexity of the network [5]. While this framework can vary according to the nature of the technology used, the general structure is quite similar. In addition to content closely linked with Digital rights management (DRM) –that is a key issue for any kind of content distribution and especially for movies - the VOD system requires software, hardware and network assets and creates new middle layers in the value chain:

- head-end video servers and platforms (encoding content

charge for their content and the greater the usage that can be expected. Currently, the premiere pay-per-view window is lagging behind the home rental window by an average of 50 days. The effect of that gap is to dampen demand for VOD and limit the price that can be charged for a VOD movie"[11].

² by bundling together telephone, Internet, and TV services

- servers, middleware)
- system management functions (control access, DRM, billing functions can be also included)
 - transport/access (transport across the network from the head-end to the DSLAM or cable hub near the consumer and connection over the “last mile” to the subscriber’s Set Top Box)
 - end-user equipment (Set Top Boxes)

The table 1 below presents a framework showing how the VOD industry is organised. For the first steps, it relies on the cinema value chain presented above. Some stakeholders have several positions in the VOD value chain by aggregating, selling VOD services directly to consumers and operating their own network (Ebiscom/FastWeb in Italy)

Table 1: VOD value chain

Activity	Function	Stakeholders involved
Content provision	produce movies, manage rights, market programs	Content providers, movie studios, film companies
Services provision	Stock films on servers (based on technologies for encoding such as digitalisation and compression) and then distribute them either through downloads (temporary) or streaming, deliver VOD for the consumer over DSL, cable. Multiple functions: <ul style="list-style-type: none"> - negotiate rights with content providers, - ensure the remuneration of the holders of the rights, - manage all the data associated with the program (language), - ensure the protection of these programs by appropriate encryption, - update the offer according to consumption parameters and contractual rights, - produce programs guides and graphic interfaces, - set up a commercial policy suited to the various types of products - manage invoicing and payment - ensure promotion of the service - negotiate agreements with physical distributors (network operators) for integration in the platform’s offer, either under the name of the end distributor, or under the service provider’s own, or using the “co-branding” system that enables each party to make the most of its contribution 	Specialised portals, content, aggregators video services providers
“Physical” transport and/or commercial distribution	<ul style="list-style-type: none"> - construct a server-client architecture dedicated to VOD (including a billing system) - provide the infrastructure and access networks to carry VOD 	Access network providers, broadband network operators, ISPs
Consumption		End-users

Source: Adapted from activities described in [10]

2.3 Issues at stake and threats

We have provided a basic level for how the VOD is organised and we have to highlight some issues it faces:

- the increasing weight of Microsoft in the film industry process with the use of Windows Media player.
- even if bandwidth and services issues are resolved, VOD will face the same problems of piracy and content protection as the music industry through freely available movies through peer-to-peer distribution over the net.
- DVDs have provided huge revenues to distributors over the last five years. Video rentals and DVD purchases account for as much as 50% of a film’s revenues. A possible risk of an online distribution is to jeopardise the revenue stream from DVDs.

3. THE MOVIESYSTEM CASE

3.1 Case presentation

Created in March 2000, Moviesystem has become the leading European VOD operator. Starting from a traditional activity of purchasing of exclusive rights for

movies and wholesale distribution of VHS cassettes to professionals, Moviesystem entered the VOD market with the acquisition of a VOD service on the Internet, NetCiné.com. Moviesystem was initially supported by Pathé and Europa Films (Luc Besson’s company), which provided funds and the VOD rights for their catalogues. Consequently supported by innovation funds such as OTC Innovation 1 and A plus Finance, Moviesystem has signed numerous agreements with content providers and now has a catalogue of 1500 films from 70 production companies. In May 2004, Moviesystem was bought by Canal +, a leading French and European Pay-TV company. Moviesystem’s three main activities in VOD business are:

- the purchase and management of exclusive rights,
- the development and operation of a specialised middleware platform,
- the commercial operation of VOD services.

Moviesystem’s middleware platform is organised around Windows Media™ 9 technology. This platform allows the addition of meta-data to protect rights, the digital coding and the encryption of films and other programmes. It also provides a full range of features, such as management of content, conditional access and

Digital Rights Management, user interface customisation, reporting and edition of statistics, and bandwidth management.

The IP-based streaming platforms can be installed in the operators' network. In order to ensure quality and speed of streaming, as well as to deal with peaks of demand, it is better to have numerous servers located as near as possible to the end users.

Moviesystem addresses the VOD market through both direct and indirect distribution. As concerns direct distribution, Moviesystem operates its own service directed at Internet users, called NetCine. NetCine's web site offers films for streaming that are grouped in various categories (such as news, comedy, drama, children, adult, etc). The film catalogue is updated every week with titles recently released on the video rental market. In order to simplify and secure transactions, Moviesystem introduced an electronic purse system with tokens. Customers can buy 5, 25 and 60 tokens, for a respective price of EUR 5, 20 and 40. A recent film is offered 9 months after its first showing for 5 tokens, that is between EUR 3.30 and 5.00 (depending on the token package concerned), a price that is comparable to the rental of videos. Once a programme has been purchased, it can be accessed for 24 hours after which the subscription expires. NetCine claims a customer base of 30,000 individuals, and direct distribution is not envisioned as the most promising channel compared to indirect distribution via ISPs or other distributors. However, it has the advantage of serving as a showcase for prospective distributors and also as a test market. It also allows the distribution of adult content, which may not be in accordance with the policy of some distributors.

As concerns indirect distribution, Moviesystem has mainly developed partnerships with ISPs³. Under each ISP's brand name, NetCiné's services are distributed by the main French ISPs, whether focused on ADSL, such as France Télécom's Wanadoo, Club Internet, 9 Télécom and AOL, or on cable TV (Noos). At the European level, Moviesystem is also present in the Netherlands with the operator KPN, and expects to extend its distribution in the short-term in Belgium and Italy. In France, the indirect distribution channel allows Moviesystem to address a potential market of 2 Million users out of a total of 3.3 Million broadband access subscribers. Moviesystem is also involved in new developments resulting from the adoption of "triple play" strategies by the main ISPs. It participates in the launch of an innovative VOD service with Monaco Telecom, where the film can be viewed directly on a TV screen, and also works with France Télécom, Free or 9 Télécom, on similar VOD services that complement their new on-line TV services.

³ A selection of the Moviesystem catalogue is also available on the web sites of FNAC, a French retailer of cultural and entertainment goods, and of Virgin.

Moviesystem is able to provide to its distributors with a complete and customised solution. Moviesystem is in charge of the updated provision of movies, the analysis of customers' habits, the customisation of the user interface, and the management of digital rights, whereas distributors are responsible for providing the suitable infrastructure, including servers located near end-users, and for billing users. The splitting of revenues is the following: rights owners get 50 % of the revenues, an average of 5 % is spent for the billing platform used, and the remaining 45 % is more or less equally shared between Moviesystem and the ISPs. Indirect distribution offers Moviesystem many advantages: a larger potential market, an extended server infrastructure, a billing system that users are already accustomed to, as well as benefiting from the attractiveness of bundled offers. However, as ISPs are keen to protect their customer base and to keep control of their billing system, Moviesystem does not have full access to customer information and lacks pricing and billing flexibility.

3.2 Case Discussion

Moviesystem's success can be attributed to the nurturing and ownership of critical resources, as well as the value created by its particular business model.

First, Moviesystem has been able to use and/or develop relational assets with both upstream and downstream partners. It was able to mobilise the relationships with producers it had developed with its original business of off line film distribution, in order to get its initial funding and most of all the VOD rights for their catalogues, and also to nurture new relationships with the main ISPs that distribute its offer. Secondly, because of its intellectual assets and its technical resources, it was able to link two different industries with different cultures, experiences and resources, thus reducing uncertainty and risks for both upstream and downstream partners.

On the one hand, firms in the media industry traditionally distributed their films via movie theatres, TV channels or video rental outlets. They were therefore not familiar with dealing interactively with end customers, such as analysing their behaviour and preferences and managing direct or indirect transactions with them. They are also not accustomed to the technical and management aspects of computing and communications systems. A firm such as Moviesystem offers them a simple solution in order to avoid these uncertainties and the risks associated. They provide their catalogue and Moviesystem manages all the various activities related to VOD, such as managing technical platforms and data associated with the programmes, processing customer information, ensuring protection of digital rights, and managing transactions with end users and distributors.

ISPs and network operators, on the other hand, lack experience in managing a catalogue of films. It requires

artistic and marketing capabilities, such as the ability to assess the commercial value of films on this complementary market, to package them, and to put them together to provide an attractive offer. From this perspective, Moviesystem benefits from its experience in the videocassette market, of which the VOD market is expected to be a partial substitute. It also necessitates more technical capabilities such as conceiving programme guides and graphic interfaces or managing digital rights. Moviesystem provides its distributors with a full service, allowing them to focus on their core activities such as managing infrastructures, customer relationships and billing.

Drawing from different theoretical perspectives⁴ [1] it is possible to assess the value creation potential of e-businesses on four interdependent dimensions, namely: efficiency, complementarities, lock-in, and novelty. We can use this framework to further evaluate the value created by Moviesystem.

Moviesystem clearly improves the efficiency of transactions. By being able to link two different industries, we have seen that it significantly reduces the uncertainty resulting from the strong information asymmetries between upstream and downstream partners, and therefore the search costs that they would have experienced without Moviesystem. It also contributes to reducing search costs for end consumers because of the provision of guides and interfaces. Furthermore, its technical platform allows opportunism to be reduced due to the efficient DRM and billing interconnection system implemented. Finally, by allowing a fully digital value chain, it simplifies transactions, increases their speed and contributes to reducing their monetary costs.

Concerning complementarities, Moviesystem does not really provide complementary outputs to customers by bundling separate products and services, as it is focused on a specific service, but a certain level of complementarity is nonetheless achieved with the creation of catalogues putting together an attractive set of various programs. By providing a digital service that fits with the services provided by ISPs, it also allows its downstream partners to benefit from complementarities.

In terms of "lock-in", referring to the level of dependence experienced by customers and partners that may prevent them from changing their supplier or partner, our opinion is more ambivalent. On the one hand, ISPs are dependent on Moviesystem because of its intellectual assets (ability to assess the value of a film, ability to bundle a catalogue of films and other items), as well as its relational assets (with the media industry). Content producers are also dependent on Moviesystem because of the trust resulting from its successful experience in

managing a DRM and billing interconnection system. However, the end consumers clearly belong to the ISPs, who manage customer relationships and billing. This will be certainly be accentuated by the bundling strategies of these ISPs, such as "triple play" strategies.

Allowing a new way of distributing movies by linking two industries and by providing an efficient support to new transactions, the business model of Moviesystem relies by definition on novelty. In fact, Moviesystem is applying to itself a process of Schumpeterian creative destruction, as its new VOD business may well destroy its traditional business of wholesale distribution of films on VHS cassettes. By being the first in France and one of the firsts in Europe, it may benefit from a competitive advantage over new players if strong learning effects, or time compression diseconomies [8], are present.

4. CONCLUSION

The Moviesystem case suggests an interesting perspective on the topic of convergence: instead of one convergent industry, the persistence of two specific industries with specific resources (Media, and Information and Communications Technology industries) that are bridged by Moviesystem. It also outlines the importance of market-based assets, as advocated in [14], composed of relational and intellectual assets. In this case, the critical intellectual assets are the understanding of both environments, including their cultures, markets, products, processes and business models. The critical relational assets are the relationships developed with upstream and downstream partners, and the resulting trust and reputation. Moreover, Moviesystem was able to leverage these assets by managing efficiently processes such as product innovation, Digital rights Management, Customer Relationship Management or billing processes.

REFERENCES

- [1] Amit, R., C. Zott, "Value creation in e-business," *Strategic Management Journal*, Vol. 22, No. 6/7, pp.493-520, 2001.
- [2] Axelsson, B., G. Easton, *Industrial Networks - A new View of Reality*, London: Routledge, Inc., 1992.
- [3] Barney, J. B., "Firm` Resources and Sustained Competitive Advantage," *Journal of Management*, 17, pp.99-120, 1991.
- [4] Berne M., "Business models for Services: the case of Video of Demand," STRAT&MARKETIC internal paper, INT, 2003.
- [5] Berne M., N. Daidj, O. Epinette, "IP Services Billing," STRAT&MARKETIC internal paper, INT 2003.
- [6] Collis D. J., P. W. Bane and S. P. Bradley (1997), "Winners and Losers. Industry Structure in the Converging World of Telecommunications, Computing, and Entertainment," in Yoffie (ed.), *Competing at the age of digital convergence*. Boston MA: Harvard Business School Press, 1997.

⁴ Value-chain analysis, Schumpeterian innovation, the resource-based view of the firm, strategic network theory, and transaction cost economics.

- [7] Debande O., G. Chetrit, The European Audiovisual Industry: An overview, European Investment Bank (EIB) Sector Papers, September 2001.
- [8] Diericks, I., K. Cool, "Asset stock accumulation and sustainability of competitive advantage," *Management Science*, Vol. 35, pp.1504-1511, 1989.
- [9] Epinette, O., O. Segard, P. Vialle, "The structuring of a value chain in order to deliver new system-based services: the case of wireless Internet for Paris airports", proceedings of the International Conference on Service Systems and Service Management (ICSSSM04), Vol.1, pp.872-876, Beijing, Jul. 19-21, 2004.
- [10] Guillou B., Online Distribution of Films, European Audiovisual Observatory, May 2004.
- [11] Isacowitz R., "VOD and the case for Content Protection", *Converge Network Digest*, November 17, 2003.
- [12] Porter M.E., *Competitive Advantage: Creating and Sustaining Superior Performance*. New York: The Free Press, 1985.
- [13] Schumpeter J.A., *Capitalism, Socialism, and Democracy*. New York: Harper, 1942.
- [14] Srivastava, R. K., L. Fahey and H. K. Christensen, "The Resource-Based View and Marketing: the Role of Market-Based Assets in Gaining Competitive Advantage," *Journal of Management*, Vol. 27, pp.777-802, 2001.
- [15] Vialle, P., *Stratégies des opérateurs de télécoms*. Paris: Hermès, 1998.
- [16] Williamson, O.E., "Transaction cost economics: the governance of contractual relations," *Journal of Law and Economics*, Vol. 22, pp.233-261, 1975.
- [17] Zhu K., "Internet-based Distribution of Digital Videos: the Economic Impacts of Digitization of the Motion Picture Industry", *Electronic Markets*, 2001, vol.11, n4, pp. 273-280.

Strategy and Risk Analysis of the ERP Projects in Three Chinese Cigarette Enterprises

Yan Zhu, J. Chen

Research Center for Contemporary Management, Tsinghua University, Beijing 100084, China
{zhuyan,chenj}@em.tsinghua.edu.cn

ABSTRACT

During the past decade, Chinese enterprises have achieved great improvement in terms of competition ability. Information techniques are playing even more important role in this period. In order to conclude principles of IT transformation processes of Chinese companies, three typical Chinese cigarettes companies have been investigated in this paper. After introducing the IT evolutions in each firm, we compared the main parameters of their ERP or similar IT projects with another foreign company. Based on the comparison, two problems are discussed further. Firstly, we use strategic grid analysis method to find what is the IT strategy of each firm and why they adopt such a strategy. The new possible IT transformation ways are discussed also. Secondly, risk management problems are another focus of this paper. Three main risk factors are integrated to judge the risk type of each ERP projects, and different risk management strategies are discussed also.

Keywords: ERP, e-business, risk analyses, strategic grid

1. INTRODUCTION

IT as a new kind of communication and innovation tools, it has had impacted enterprises management processes greatly. A lot of new management concepts, such as ERP (enterprise resource planning), CRM (customer relation management), SEM (strategy enterprise management), have been put forward to improve the core competition ability of enterprises. Especially for the globalizing enterprises, they can not survive without modern IT-based management methods. During the past decade, the pioneers have invested a lot in combining IT with traditional management methods.

Chinese enterprises are playing a more important role than ever in the global economic system now. As the most potential market, it is very important to analyze the IT-based business models of Chinese enterprises. These kind of research works will help Chinese companies to understand how to rapidly adapt the IT transformation process. At the same time, it will be also useful for the foreign enterprises to know the mutually impact between oriental culture and information technology.

The purpose of this paper is to survey several Chinese cigarettes companies, compare the different IT transformation ways of each company, and discuss their e-business strategies and risk management methods during the IT absorption processes.

By summarizing the different IT evolution of these enterprises, and concluding why they adopt current IT strategies and how they established these strategies, the effect of their IT strategies, especially the impact on the profit, human resources, and productivity is concluded.

Based on the conclusion, we use strategic grid method to further discuss IT strategies of each company, and put forward new IT transformation ways correspondingly. In terms of risk management of ERP projects, we list three important factors that affect risk management strategies, and analyze the risk of each company theoretically.

2. IT EVOLUTION IN EACH FIRM

2.1 SAP R/3 in Yuxi Hongta Tobacco (Group) Co., Ltd.

In order to improve the group's competition ability, Yuxi Hongta has explored the IT-based manufacturing and management systems for nearly ten years. In 1993, the group began to invest on the automatic manufacture systems as well as some management systems year by year, by the end of 1998, more than 300 computers were running for line-control as well as management. In addition, the group also set up 10 ALPHA series mini-mainframe computers, 4 PC servers, 27 SGI Graphic workstations. In 1995, the FDDI project had been finished, and the local computer network was established in 1996 which had been upgraded to ATM in 2000. There were 48 application software systems in different sections of the group before 2001.

In 1998, the group started to contact ERP concepts, and compared different implementation strategies of ERP, such as applying advanced ERP software package, applying the manufacture-based medium ERP system, developing the ERP system by themselves. After serious consideration, they decided to adopt the first scheme,

and in 2000, the contract had been signed with SAP AG (China), and the ERP project began on Jan. 12, 2001, finished on Feb. 12, 2002.

On Feb. 12, 2002, the ERP system of Yuxi was online successfully. The improvement is dramatic, just as Mr. Zeng, the CIO of the group, mentioned, the enterprise change its operation way from the deep inside, the mainly improvements include:

- 1) Change of enterprise strategy. Close-loop business processes had been set up. So the leader group can get information about manufacture and market immediately and make decision more accurately. In addition, the enterprise has formed an IT-based strategy, which is mainly about how to convert data into knowledge of the enterprise.
- 2) Logistics had been optimized. Mr. Zhang, the chief manager of Hongta material company, said, ".....in the ERP system, Inventory management and finance management are integrated, once materials were shifted, the system will update the account automatically. At the same time, system can generate online inventory reports; it is very convenient for query. Because of the implementation of ERP system, the inventory cost had been reduced 0.2~0.5 billion Yuan (approximately 25~50 million USD). The system also enabled the total cost control, it is estimated that the production cost will reduce 50~200 million Yuan (approximately 6~25 million USD).....".

At the same time, the loading efficiency had been improved also, loading ability had been changed from 8,000 cases per day to 11,000 cases per day.

- 3) Change of supply chain management. In the ERP system, information among the whole supply chain (suppliers, purchasers, manufacturers, wholesalers and retailers) was integrated, managers can learn the whole status of the supply chain, and make decision and coordination accurately and immediately. Especially for the suppliers management system, because of the RFQ (Require for Quote) and PO management system, managers can compare different suppliers, and greatly reduced the purchase cost.
- 4) Change of finance management. The budget was made and evaluated according to the standard cost system in ERP system. So, any finance problems can be found by querying the accounting system, and the reason and responsibility of these problems can be addressed easily also. The new characteristic of the finance system is integration; module FI (Finance) and CO (Control) are fully integrated with MM (Material Management), SD (Sales and Distribution) and PP (Product Planning), which make the whole system is consistent and efficient, and provide the manager accurate decision information.

2.2 Oracle Applications in Baisha Group

IT utilization in Baisha group can retrospect to the year of 1992; the former CEO realized that information technology will enforce the manufacturing ability, so they decided to invest on IT infrastructure. By the year of 2000, 60 million Yuan (approximately 7.5 million USD) had been spent on IT projects, the basic IT infrastructure had been established and a lot of application software systems were running in different departments.

Although the group had spent so much money, the CEO recognized that their IT expenditure was not successful until 2000. Especially for the manufacturing system, the group spent more than 2 years to implement MRPII (manufacturing resource planning), but the results were not satisfied. By that time, the leader group of Baisha analyzed experiences of their IT expenditure, and some defects were found out.

According to the analysis above, the group decided to implement totally IT-based management reform project. They called it BPR & ERP project. It was quite hard project, because the reform process will change the balance of the company, in case of the project was failure, the company may be disorder.

After carefully comparing different ERP software vendors, the group decided to adopt Oracle Applications 11i. The flexibility of Oracle Applications permits the group to design new business processes easily and quickly. And the implementing cost was relatively low also.

The typical characteristic of the implementation is that the group made great effort to do BPR Project first, and then implemented ERP software system. On Jun. 25, 2001, the BPR project had been completed; on Nov. 5, 2001 ERP system was launched.

Talking about the improvements, Miss LU Ping, the CEO of the group, said "..... the remarkable improvement is that we has rebuilt our Baisha culture, and the business processes have been optimized also, so the Baisha group become stronger and agile.....". To summarize, the achievement in detail includes:

- 1) Formed new enterprise culture. All employees have had a clear and same goal, and put great enthusiasm in their work now. They have learned the new position of themselves in the new business process, and were encouraged by the perspective of the enterprise. The new company culture, "long-lived brand, long-lived enterprise and long-lived Baisha people", has been carried into each job position.
- 2) Optimized business processes. During the 51 core business processes, the non-value-added processes had been eliminated, the value-added parts had been

reinforced. And the manufacturing recycling time had been shortened simultaneously.

3) Organization reform. According to the new strategy and the principle of lean and intensive of the new enterprise, the group re-designed their organization architecture, set up a new structure based on the new business process.

4) Improved manufacture and management efficiency. The ERP system had been successfully implemented, and the optimized business processes were solidified in the system. So the company is run more efficiently than before, and the reward was remarkable. Such as, accuracy of financial plan had been improved 15%; accessory material inventory cost had been reduced from 100 million Yuan to 50 million Yuan; delivery cycle (from moving out inventory to loading) had been shortened from 12 hours to 10 minutes.

2.3 Develop ERP by themselves: Honghe Group

Before the year of 1997, although Honghe Group never considered IT-based business strategy, their automatic manufacturing system had been an important infrastructure of the company. Leading by Mr Qiu, the CEO of the group, the company has established some special but efficient business processes.

When they began to consider the IT strategy in 1997, after comparing a lot of ERP software packages, they found that it was quite difficult to choose one kind of software which can be compatible with the automatic manufacturing system as well as the existing effective business processes. It seemed that the best way is to develop the integration system by themselves.

To make such a decision was quite hard, because the company had not enough information technical teams. But fortunately, during the past 5 years, the great partner—Kunmin Ship Company—had built a strong IT team which can accomplish the software development tasks.

The core of the project is how to design business processes. From 1998, Mr. Qiu, served as both CEO and CIO, lead the project himself, and attended to design each business process. It took 3 years to accomplish this work, during the 3 years, Mr. Qiu never left Yunnan province, he put all his efforts on the redesign of the processes. It is very rare that the CEO will design the processes by himself during an IT project, but Mr. Qiu made it.

Talking about the scheme of the IT project, Mr. Qiu is always exciting, “.....the project is a system, it must embody the characteristics of how to operate cigarettes manufacturing systems, so it is the operating software of the enterprise. The scheme must be efficient and optimized, the organization structure, business processes

and the operating rules must be totally designed in the scheme. And at the same time, the company should rebuild it's IT-based cultures as well.....”

The business processes scheme includes 8 modules, that is: office support, finance, manufacture, sales and distribution, raw materials management, auxiliary material management, equipment management and R&D. All of these modules can be fractionized into more than 11,000 sub-modules.

During the three years design process, the company reformed their organization structure accordingly as long as opportunity was proper. Before the HY-CIMS project began in June 2002, they have established firm fundamental of both organization and knowledge.

In June 2002, the company shifted their scheme to Kunming Ship Company (KS), and the KS finished the blueprint of HY-CIMS technically on Jun.18, 2002. From Jun. 2002 to Aug. 2002, KS accomplished the detail design of each module. Finance module had been developed at the same time, and the first training about new finance system was carried out simultaneously.

By the end of Oct. 2002, modules had been online continually, material management, manufacturing, office support, personnel management and equipment management modules were developed as well as trained. In Jan. 2003, the whole system had been integrated and tested; HY-CIMS had been accomplished.

The HY-CIMS project brought great effect on the company, although the investment is no more than 10 million Yuan (approximately 1.2 million USD). The direct benefits are the costs have been reduced and the efficiency improved. According to the preliminary estimation, HY-CIMS has improved the yield of the company by 3%, this means the company can get another approximately 75 million Yuan (9 million USD) each year.

3. STRATEGIC COMPARISON OF IT EVOLUTIONS

In table 1, we conclude the basic IT transformation process of each company.

It has been suggested that to fully realize the competitive advantages created by the strategic use of IT, organizations must have appropriate IT management strategies that fit with their culture, technology status, and business strategy.

This contingent nature of IT management has drawn much attention in the IT literature^[1, 2, 3]. Given the importance of a proper alignment of IT management processes with the organizational role of IT, organizations need to have a clear and systematic understanding of the strategic relevance and impact of

their IT, and then develop IT management strategies appropriate to that role.

A framework for delineating the differences in the strategic relevance and impact of IT was proposed by Applegate, McFarlan and McKenney^[1]. This framework

identifies four different IT roles which are referred to as categories of strategic relevance and impact or, simply, as the strategic grid. The grid is a useful framework for understanding the positioning of an organization's IT, and as the basis for designing appropriate planning and control systems for IT management.

Table 1 Comparison of IT Transformation Process

		Hongta	Changsha	Honghe	Dow Corning ^[4]
Basic Information	Yield (2002)	220	90	90	
	Income (Billion USD)	4	0.9	0.9	2.5 ¹
	Employee	4500	2600	1800	9000 ¹
	5 year Income Aver. Increase	15%	20%	20%	~4%
Strategic role for IT	Begin Date	1993	1992	1996	1996
	Focus	Business Change	Business Change	Business Change	Business Change
	Strategy Leader	VP	President	President	President
ERP	Combine with BPR?	Half	Yes	Yes but passive	Yes
	Project Initiated	2001.2	2001.1	1999.1	1996
	Completed	2002.1	2001.11	2003.1	1999
	Period(month)	11	10	48	36
ERP	Expenditure (million USD)	10	2	3	<100
	Project Manager Consultant	IT manager RenTong Cons.	President HAN Consulting	President None	VP/CIO None
	Software Vendor	SAP mySAP.com	Oracle	Developed	SAP R/3
Project Benefits	Hierarchy Reform	Done in 2003.4	Same time	Little	Yes
	Reputation	Improve a lot	Medium	Medium	Improve
	Enterprise Culture	Improve	Rebuilt	Little Impact	Coming Impact
	Profit Increase (EST)	5%	3%	5%	N/A
	Inventory	-20%	-50%	-30%	
	Employee	-20%	-30%	-	
Others	IT % Income	~0.5	~0.1	~0.2	~4.5
	IT Reporting Influence	Low	High	Medium	High

Note 1: Data in 1998

* Yield unit: million cases

Applegate, L. M., McFarlan, F. W., and McKenney, J. L.,^[1] describe two (future and current) aspects of strategic relevance and impact of IT and are referred to as the future and current portfolios, respectively. The strategic relevance and impact of IT for a firm can be positioned anywhere on a continuum ranging from low impact and relevance to high impact and relevance along each of these dimensions of the grid. Applegate, L. M., McFarlan, F. W., and McKenney, J. L., propose that understanding an organization's position along these continuums is critical to developing appropriate IT management strategies. The combination of future/current portfolio with high/low impact presents four types of IT environments and is represented by the four cells (strategic, turnaround, factory and support

cells) of the strategic grid.

Strategic (Cell 1): Organizations in this cell are critically dependent on the smooth functioning of the IT activity. This dependence relates to existing systems as well as to applications under development, i.e. to both the current and future portfolios.

Turnaround (Cell 2): While existing applications may provide operational support for organizations in this cell, they may not be critical to the organization's current operations. However, the impact of the applications under development is considered vital to achieving the organization's strategic objectives.

Factory (Cell 3): These firms are heavily dependent on

their existing, day-to-day IT operations and even a temporary disruption in service could have severe operational consequences. However, the applications under development for the future are not critical to the firm's ability to compete successfully and are mostly maintenance projects.

Support (Cell14): In these firms, neither current nor future IT activities are critical to the smooth operation of the firm. IT is viewed as playing the traditional back-office-support role without any potential for dramatically altering an organization's current or future operations and/or direction.

According to the characteristics of each firm, they can be located in different cells.

Hongta: this firm is in the strategic cell, because they have completely established an ERP system, and totally considered the business strategy based on IT strategy.

Baisha: this firm is in the transferring process from factory cell to strategic cell. The firm has reformed the business process and the integration system which has already been online and under development will greatly impact the business strategy.

Honghe: is in the process from turnaround cell to strategic cell. Although they also considered the IT-based business strategy, the firm still need integration of manufacturing system and management system.

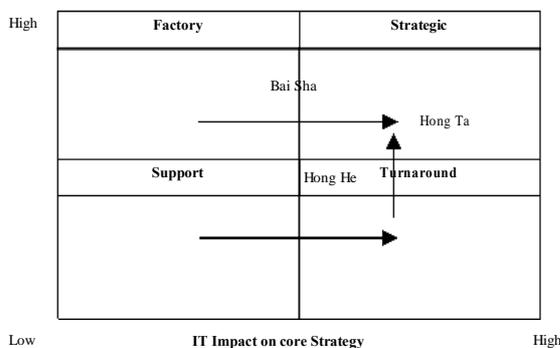


Figure 1 IT Strategic Grid of Each Firm

4. RISK MANAGEMENT METHODS COMPARISON

IT projects are high risk projects, it need careful consideration of the risk management. During the

process of implementing the IT projects, all of the three firms considered the projects risks in different way, and adopted different way to avoid risks.

In a recent literature published by Prof. Cyrus Gibson [5], the IT project risks can be managed through a process from risk reorganization, risk assessment to risk reduction and cope with remaining risk. The methodology can be described in Figure 2. According to leadership, employee's perspective and project urgency, different implementation strategy should be adopted to avoid risks. For example, if the leadership is positive, the employee's perspective is negative but the project urgency is positive, then the implementation strategy should be guided evolution.

The IT projects implementation processes of three cigarettes firms can be assessed in table 2. That is, for Hongta company, the leadership is positive but the employee's perspective and project urgency are negative, so the theoretical implementation strategy is top-down coordinate. In Baisha, all of the three factors are positive, so the theoretical implementation strategy is big-bang. In Honghe, the leadership is positive, employee's perspective is positive, the project urgency is negative, so the theoretical implementation strategy is improvisation.

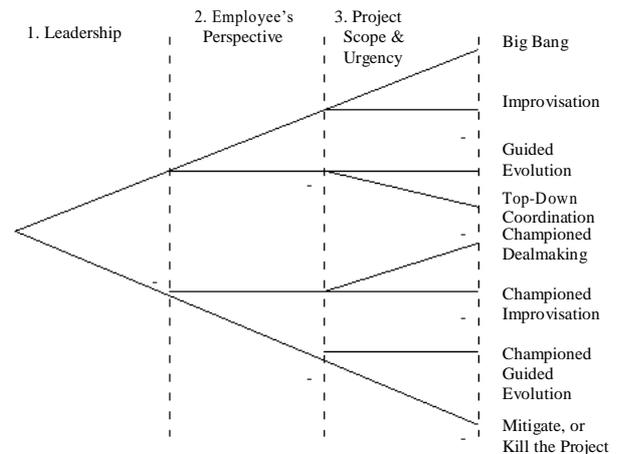


Figure 2 Project Risk Assessments

In order to achieve the actual IT projects risk management strategy of each firm, we talked with three levels of each firm, including CEO and CIO, middle manager and project team member. According to the description of how they organize the project, the actual IT project risk management strategies have been concluded and depicted in table 2.

Table 2 IT Projects Management Strategy

	Leadership	Employee's Perspective	Project Scope & Urgency	Actual PM Strategy	Theoretical PM Strategy
Hongta	+	-	-	Top-Down Coordination	Top-Down Coordination
Changsha	+	+	+	Big Bang	Big Bang
Honghe	+	+	-	Improvisation	Improvisation

5. CONCLUSIONS

According to the comparison in chapter 3, we can reach following conclusions about the IT transformation processes of each firm.

1) All of the firms have invested on IT, but they are in different stage. each of these firms have found the importance of IT, and want to improve their management as well as manufacturing system using new information technologies. But because the background and geography are quite different, they are in different IT absorption stage now.

2) Because of lack of IT strategy originally, these firms can jump from one period to another. Except for Hongta, other companies are planning to adopt the jumping strategy. That is, they hope that they can jump from lower stage directly to the highest stage. Technically it is possible, but considering the enterprise culture and management level, it is quite hard to omit one or two stages. The best way is to form a IT-based business strategy and implementing IT-related reform continuously.

3) IT does matter in Chinese enterprises. Enterprises in China are facing a great chance; they must set up new business architectures which will be competitive with other companies, especially the foreign companies. So IT infrastructures are still a problem for most of Chinese companies, and the new IT-based business processes still need to be developed. In order to enhance competitive advantage, the IT-based business strategies of each company should be set up as soon as possible.

4) Successful implementation experiences are that all of the three companies considered their enterprise culture during the ERP implementation process, and adopted different implementing methods accordingly to avoid risk. According to the project risk management theory, the actual IT projects management method is consistent with the theoretical expectation as long as the project is successful.

The works in this paper are still limited, because there are so broad topics considering the IT innovation in Chinese enterprises. Further research works will focus on IT risk analysis of Chinese state-owned large enterprises. With the rapid development of Chinese enterprises, IT will get more and more attention as well as expenditures. It is a serious problem about how to manage the IT projects risk. Although we introduce a new method in this paper, further research must be done

considering the concrete culture and diversity implementation strategies. The detail information about IT project management in each company will be collected in the next step, and we will try to find the basic principles of the risk management of IT project.

In addition, it is important to judge whether an IT project is valuable or not. There are several methods to judge the IT project value^[6]; one of these methods is using the evaluation index. But the index is quite different because of different culture background. According to the successful experience of Chinese companies, the evaluation index should be studied carefully, and a standard of evaluating IT projects should be given.

IT strategy is so close to business strategy. With the support of IT and IT-related new strategy analysis methods, how to establish competitive business strategy is very urgent problem. Some new characteristics and improvement should be made to use the strategic grid and other methods.

ACKNOWLEDGEMENT

This paper is partly supported by the National Natural Science Foundation of China (70231010, 70101008).

REFERENCES

- [1] Applegate, L., F. W. McFarlan and J. L. McKenney. *Corporate Information Systems Management*. Chicago, IL: Irwin, 1996
- [2] Applegate L., Robert D. Austin, F. Warren McFarlan, *Corporate Information Strategy and Management*, sixth edition, McGraw-Hill Higher Education, 2003
- [3] Peter Weill, Marianne Broadbent, *Leveraging The New Infrastructure: How Market Leaders Capitalize on Information Technology*, Harvard Business Scholl Press, 1998
- [4] Jeanne W. Ross, *Dow Corning Corporation: Business Processes and Information Technology*, Sloan School of Management Case, April 1997
- [5] Cyrus F. Gibson, *IT-Enabled Business Change: An Approach to Understanding and Managing Risk*, MIS Quarterly Executive, Vol.2, No.2, September 2003.
- [6] Audrey Apfel, Michael Smith, *Gartner Group Report: TVO Methodology: Valuing IT Investments via the Gartner Business Performance Framework*, 03 March 2003, Document Type: Strategic Analysis Report, Note Number: R-19-1910

Towards a Knowledge Based Society: A Case Study on Entrepreneurship in Malaysia

Goh Hong Wee, Goh Pek Chen

Faculty of Business and Law, Multimedia University, Malaysia
{hwgoh, pcgoh}@mmu.edu.my

ABSTRACT

This paper as a whole will discuss the entrepreneurship development in Malaysia. The 21st century promises a world of challenges and opportunities to all the nations in the world. Malaysia has launched the Multimedia Super Corridor (MSC) to spearhead Malaysia's challenges in the new millennium. In line with Malaysia's objective to produce world class Information Technology (IT) –savvy-working forces, a University was born in center of Cyberjaya called Multimedia University (MMU). MMU is playing a pivotal role in the national entrepreneurship development in Malaysia, similar to that of Massachusetts Institute of Technology (MIT) and Stanford University in United States. In Malaysia, various business plans are held throughout the year providing university students a platform to demonstrate their entrepreneur skills. Besides that, many venture capitals sources are available in Malaysia for potential technopreneurs to seek funding and advise. Successful Malaysian entrepreneurs such as Mark Chang (CEO of Jobstreet.com) and Tony Fernandes (CEO of AirAsia) can be the inspiration for many Malaysia's aspirants.

Keywords: *Entrepreneurship, Knowledge, Society, Technopreneur*

The Knowledge Era has reached; most of countries in the world have tried to work towards a Knowledge based Society. Malaysia on the other hands, do not lose out in the race of the globalisation. As a major step in building up a Knowledge Based Society, Malaysia has established a major information technology (IT) project called Multimedia Super Corridor (MSC) in 1996 to prepare its people for the knowledge revolution in the new century.

Under the MSC project, the development of entrepreneurship is given a high priority. A MSC technopreneurship development flagship was then established in 2001. The core objectives of the flagships are to facilitate the developments of technopreneurs, start –ups and existing information communication and technology (ICT) companies and to assist and facilitate the growth of ICT small and medium enterprises (SMEs) into world-class companies. The flagship aims to drive entrepreneurship in Information and Communications Technology (ICT) and other high-tech areas. Not only that the Technopreneur Development Flagship aims to strengthen the ICT, Biotechnology and life science industries in Malaysia by spawning a critical mass of entrepreneurs and nurturing world-class companies.

In order to successfully steer the country into the Information Age, the flagship not only pull together all efforts at Federal and State level, but also the private sector to boost the development of SMEs throughout the country. It is because if the flagship is too concentrated at federal level only, the problem of digital divide in Malaysia will be worsened.

With the co-operation of State governments as well as support from the private sector, Technopreneur Development Flagship aim to expand this network through the cultivation of new incubators, with at least

one incubator in each state. To encourage the setting-up of incubators, operators will receive MSC status and enjoy the incentives outlined in the Bill of Guarantees. Companies based in these incubators will also benefit from some of these incentives

In order to prepare a conducive environment for Malaysia's aspirants, a National incubator network has been established. Under the National incubator network, there are currently 9 incubators in Malaysia namely MSC Central Incubator, Kulim Technology Park Corporation Sdn Bhd , BT Multimedia Sdn Bhd, N2N Venture Solutions Sdn Bhd, YTL e-Solutions Bhd, Makmal.com Sdn Bhd, Usains Tech Services Sdn Bhd, BioEnterprise Asia Sdn Bhd and Bureau of Innovation & Consultancy (BIP), Universiti Teknologi Malaysia. Attractive incentives are given to incubator operators such as freedom of ownerships, income tax exemption, intellectual property protection, competitive telecommunication tariff and many more.

2 prototyping labs focused on enhancing technology skills and building Malaysia's human capital in the areas of advanced software development, embedded systems, intelligent systems and wireless & mobility was established too namely Java Technopreneur Development Centre (JTDC) in partnership with SUN Microsystems (SUN) and Multimedia University and .NET Technopreneur Development Centre (NTDC) in partnership with Microsoft. Under the MSC Technopreneur Development Flagship programme, 9 teams of Microsoft .NET software developers have been selected out of 90 to help to turn their ideas into viable businesses.

At university level, the work of promoting entrepreneurship has been quite aggressive. Multimedia University, a university, which has been given the role of

producing entrepreneurs in Malaysia, has established the Centre for Commercialisation and Technopreneur Development [CCTD] for both the commercialisation of Multimedia University's research and also the development of technopreneurs among the university community. CCTD is seen to provide the link between academia and industry for the development of commercially viable technologies and is committed to building beneficial partnerships between academia and the commercial sector.

Currently there are a few starts –up companies set up by the Multimedia University students namely RiteUni Dot COM, Xirien Technologies, E-Transact and many others. Multimedia University latest start-up will be Roommart. Roommart was founded by an undergraduate of Multimedia University. Roommart is Malaysia's 1st online roommate finder. It is a portal, which helps students from higher learning institutions to find rooms, roommates and houses. Roommart define a new way of finding accommodation through Internet as most of the students today are avid Internet users.

In Malaysia, university students have a lot of business planning competition opportunities such as HSBC Young IT Entrepreneur Awards to serve as a platform for them to demonstrate their business writing skills. The 2004 HSBC Young IT Entrepreneur Awards has successfully attracted 250 proposals for the competition showing that entrepreneur competition like this are well received in Malaysia. University Cradle Investment Programme (U-CIP) is also being introduced to facilitate university research outputs into marketable products/services, as a first step in the commercialisation roadmap.

Many venture capitals sources are available in Malaysia too for technopreneur to seek funding such as Malaysia Venture Capital Management Sdn Bhd and MSC venture

capital corporation Sdn Bhd. The setting up of MALAYSIAN EXCHANGE OF SECURITIES DEALING & AUTOMATED QUOTATION BHD (MESDAQ) which is a stock exchange targeted to the growth of technology companies has creates a new vista for companies to seek for potential investors and vice versa. Through a listing on MESDAQ, companies can now have access to public funds. With the help of some organisations such as Technopreneurships Association Of Malaysia (TEAM), which acts as a think tank for technopreneurs in Malaysia, the future of entrepreneurship really look good in Malaysia.

For a developing country like Malaysia, it is very important for the nation to have its own successful entrepreneur who can become a fine example for the youth to follow. In Malaysia, the success story of Mark Chang (CEO of Jobstreet.com) and Tony Fernandes (CEO of AirAsia) has boost up the confidence of many young Malaysians and encourage them to start their own business and become an entrepreneur.

REFERENCES

- [1] Multimedia Development Corporation Official website URL : <http://www.mdc.com.my/>
- [2] Multimedia University Official website URL : <http://www.mmu.edu.my>
- [3] MSC Technopreneur Development Flagship Official website URL : <http://www.technopreneurs.net.my/>
- [4] Technopreneurs Association of Malaysia Official website URL : <http://www.team.net.my/>
- [5] Roommart Official website, URL <http://www.roommart.com.my/>
- [6] Mesdaq Official website, URL : <http://www.mesdaq.com/>

Analysis of Trademarks Act of Thailand and a Case Study: Conflict of Domain Name and Trademarks

Settapong Malisuwan

Associate Professor, College of Business Administration, Touro University International, USA.
Associate Judge of the Central Intellectual Property and International Trade Court, Thailand
settapong.m@cmmu.net

ABSTRACT

In the era of Information Technology, Internet plays an important role and becomes worldwide infrastructure of global communication. Domain Name is the key to reach the purpose derivable Website on the WWW. Confusion and conflict arise when words and names of Domain Name become low in resource and somehow have the same or similar identification among two or more parties. This paper provides a brief background on the interface between domain names and trademark. The paper presents examples of domain name disputes and analyzes some important aspects of legal principles of the Trademarks Act (No.2) of B.E. 2543 (2000). The paper also explains how the Act can protect the trademark and domain name rights. Finally, the roles of the Central Intellectual Property and International Trade Court in enforcing the Intellectual Property Rights in trademark and domain name protections are discussed including a case study. This paper will be a valuable contribution towards IT society and e-business sector.

Keywords: Domain Name, Trademark, Cyber-squatters, Intellectual Property, and Law.

1. INTRODUCTION

With the increase popularity of the Internet, e-commerce has become an increasingly important role for business and the value of domain names has increased accordingly. Internet users rely on domain names, which take the form of memorable and sometimes catchy words, to stand in their place. The uniqueness requirement creates an exclusivity that has important economic consequences. Domain names are interesting and controversial because they are meaningful. The semantic dimension allows domain names to act not only as unique addresses, but also as brand names. [1].

Companies and organizations want their websites to be easily accessible to potential consumers. They desire a domain name, which matches their corporate name or trademark, preferably in the .com top-level domain. As domain names have, in many ways, become the equivalent of a company name and address for Internet commerce, there has been increasing competition for the limited number of available domain names. Domain name trading has, therefore, resulted from this competition [2].

For examples, in July 1998, Compaq Computer Corp paid \$3.35 million for rights to "altavista.com", purchased from a California start-up called Alta Vista Technology. In late 1998, VA Software (then VA Research) bought the rights to the "linux.com" domain name for more than \$1 million. In 1999, eCompanies paid \$7,500,000 to purchase rights for "business.com" [2]. In April of 1999, "drugs.com" sold at auction for \$823,456 and "wallstreet.com" attracted a cool \$1.03M. In August 2000, the domain name "cyberworks.com" was sold for \$1,000,000. In February

1999, the Houston entrepreneur sold the domain name "eflowers.com" to the owner of Florida-based Flowers Direct for \$1M [2]. In 2000, Bank of America has purchased the domain name "loans.com" for US\$3 million. It bought the rights to the domain name from "GreatDomains.com" on January 28. The seller was San Jose computer consultant Marcelo Siero, who registered the domain name in September 1994 for creation of an e-commerce application but later decided to sell it, according to "GreatDomains.com" [3]. "Wines.com" were bought for \$3M. "Websites.com" was auctioned off for \$0.97M by Great Domains (www.greatdomains.com) and now the website provides Internet and web hosting services to businesses. "Telephone.com" was acquired for \$1.75M, while "bingo.com" sold for \$1.1M [4]. Barnes and Noble bought "books.com" for \$1M. In a recent example, the government of New Zealand paid US\$500,000 to U.S.-based Virtual Countries to obtain rights to the domain name "newzealand.com" [5].

In Thailand, e-commerce has come to people's attention a great deal when MWEB (Thailand) Ltd purchased the website "sanook.com" at a very high price from an individual, Mr Porameth Minsiri, a webmaster who subsequently became a millionaire and a very popular person [6]. In 1997, Mr. Adam Stanhope learned the domain name "bangkok.com" was for sale. He negotiated to buy the domain in installments. Fortunately, the Asian currency crisis hit and he ended up only paying \$4,200 for the domain name. The website "bangkok.com" sold 100 year-long subscriptions at \$35 each. They offered free email, added content and sold advertising to those interested in reaching people who wanted to do business with the Thailand market. In addition, they offered free Web site hosting for sites about Thailand. In late 1998, the

owners had created a significant virtual community who used their chat services, Web hosting and email capabilities [7]. In January 2000, Nation Multimedia Group Pcl, which published Thailand's largest business daily newspaper and the second rank of English-language newspapers, paid less than \$7 million for the Internet address "Thailand.com". It didn't identify the seller [8].

2. ICANN AND THE DOMAIN NAME SYSTEM

In 1995, World Intellectual Property Organization (WIPO) was requested by many member countries to find ways to solve disputes on domain name. In 1998, the Internet Corporation for Assigned Names and Numbers (ICANN), is an independent, non-profit organization which was created to oversee the assignment of names and numbering to ensure assignments are made on a global level with participation of all relevant stakeholders representing all of the world's regions. ICANN is a consensus development body for the global Internet community and its focus is the development of consensus policies relating to the single authoritative root and the Domain Name System (DNS). These policies include those that allow the orderly introduction of new top level domains (TLDs) [9].

In 1999, WIPO came up with suggestion to ICANN to deal with administering Domain Name and its conflicts. And in June 2000, WIPO reconsiders its suggestion and manages to put forward concrete conclusion to clarify all problematic matters including cause of confusion or unfair utilizing [10].

Due to low resource problem, ICAAN has tried to increase the number of top-level domains (TLDs), but many companies are worried that cyber-squatters will try to register a company's trademark or name in these new TLDs and then try to sell it to the owner. In 2002, ICANN announced the seven new domain names such as dot-biz, dot-info, dot-aero, dot-museum, dot-name, dot-pro and dot-coop. However, these new TLDs will be restricted. Dot-biz will be restricted to businesses; dot-aero to airline related ventures; dot-museum to museums; dot-name to personal names; dot-pro to lawyers, doctors and accountants; and dot-coop to cooperatives. For recommendation, to avoid cyber-squatting, the domain-name owners should set up pre-registration processes to allow trademark holders to claim or reserve domain names on a first-come, first-served basis [11].

Recently, more than 33,000,000 domain names have been registered, including tens of thousands of domain names that infringe on trademarks and service marks [12].

3. CYBER-SQUATTERING IN DOMAIN NAME SYSTEM

"Cyber-squatters" are defined as those who quickly

register someone else's trade name as a domain name with InterNIC, the Virginia-based central domain name registry service [12]. Generally, however, if someone who lacks a legitimate claim registers a domain name with the intent to sell the name, prevent the trademark holder from gaining access to the name, or divert traffic, this activity will be considered cyber-squatting. Despite this strong trend against cyber-squatters, new instances of cyber-squatting continue to arise. For instance, Intel filed suit against the registrant of "pentium2.com" which leads to a pornographic web site [12].

In many cases, an owner of trademark rights can find that other persons have registered as domain names that are the same as or similar to the owner's trade mark. There is also a different kind of conflict, which can apply where an owner has trademark rights but is not permitted to register it where domain name policy prevents registration of names such as geographic names and generic/descriptive names.

The trademark owner's trade mark may be very well known around the world in its commercial context but may be a name used by others in different commercial contexts. An example of this might be the trademark McDONALDS, which is registered and well known in many countries in connection with convenience food stores but, as a common surname, may be used by many others in connection with a range of business and other activities. The trademark APPLE in connection with computers may be an example of such a name. A geographic name example may be OXFORD for publishing.

4. DOMAIN NAME PROBLEMS IN THAILAND

A major problem arising out of e-commerce worldwide, including Thailand, is the one relating to domain name registration, especially, the problem of cyber-squatters who take other person's name, trademark, service mark or trade name to register in its own name. Therefore, this section discusses the domain name problems in Thailand. The followings are examples of domain name disputes in Thailand.

Bangkok Publishing Public Co Ltd or "Bangkok Post" has used trademark "bangkokpost" with newspaper and used domain name "bangkokpost.co.th" registered with the Thai Accredited Domain Name Registrar (THNIC). The company tried to register another domain name with Network Solutions Inc (NSI) for domain name "bangkokpost.com", but found that Solberg PM&P AB, a corporation in the United States, has already registered the "bangkokpost.com". This causes Bangkokpost to use domain name "bankokpost.net" with its website for a while. However, finally, Bangkokpost has bought the domain name "bangkokpost.com" to avoid disputes [6].

In August 1997, during "Amazing Thailand" campaign period, a Canadian tried to sell his domain names

“amazingthailand.com” and “amazingthailand.org” to the Thai Government in the amount of \$2.3m. However, the Thai government refused to purchase the both domain names and decided to use domain name “tourismthailand.org” [6].

In October 2000, a Thai cyber-squatter sent an email to many well-known websites such as “mweb.co.th”, “Thaiadclick.com”, and etc. He threatened the companies that if the companies refused to purchase his domain names which are similar to mentioned domain names, he would use his domain names with pornography pictures to dilute value of the dot com companies’ trade names. Finally, he was arrested and prosecuted for extortion and blackmail, in accordance with the Thai Penal Code [6].

5. THAILAND’S TRADEMARKS ACT: ANALYSIS

Thailand is required to implement the intellectual property laws into compliance with the current WTO Agreement regarding Trade Related Intellectual Property Rights (TRIPS Agreement).

Thailand’s Trademark Act was approved by Parliament earlier this year, was published in the Royal Thai Gazette on April 1, 2000 and, became fully effective on June 30, 2000. The most recent Trademark Act is referred to as the Trademarks Act B.E. 2534 (1991) as amended by the Trademarks Act (No.2) of B.E. 2543 (2000).

In this section, the Trademarks Act (No.2) of B.E. 2543 (2000) is analyzed and explained. However, this study focuses only on important parts of the Act related to the Trademark and Domain name issues [13].

The Act contains several significant changes [14]. These changes will make the application and registration process easier for many trademark applicants. The significant changes can be analyzed as follows:

- (1) Under Section 4 of the former Act, a mark was defined as a “photograph, drawing, device, logo, name, word, letter, numeral, signature, or any combination thereof but not including industrial designs under the law of patents.”

In section 4 of the new Act, “mark” means a photograph, drawing, device, brand, name, word, letter, manual, signature, combinations of colors, shape or configuration of an object or any one or combination thereof.

Analysis: In the new Act, the definition of a mark is expanded to include “phrases, color combinations and shape or configuration of goods.” This expanded definition means that consumers are now more sophisticated in distinguish the goods or services of one proprietor from another than before. For the section 4 of the Act, it shows that the Thai law

tries to enforce and protect the owner rights on Trademark.

- (2) Section 7 (2) of the former Act can be interpreted that the name of a juristic person could only be registered as a mark if it was “represented in a special manner” and, not a geographical name.

In section 7 (2) of the new act, it can be explained that one may now register a juristic name as a mark in block letters, provided it is not the *full* name, nor descriptive. This means that companies can omit “Inc.,” “Co.,” Ltd.,” “Corporation” or other business entity designation from the representation of their mark [14].

Analysis: In section 7 (2) of the former Act, the Act implies that the name of a juristic person in mere block letters could not be registered as a mark. Therefore, it can be seen that brick-and-click companies cannot protect their trademarks in term of domain names from cyber-squatters. Luckily, for the new Trademark Act in section 7 (2), the Act helps to solve this problem. It will be of great benefit to many companies, especially for the new dot com companies seeking trademark protection for domain names.

The analysis above shows that the amended Trademark Act, the Trademarks Act (No.2) of B.E. 2543 (2000), is a very positive step forward for Thailand, setting new standards for the e-business sector to protect their Domain name and trademark rights. An example of trademark and domain name case will be discussed in next section.

6. ROLES OF THE CENTRAL INTELLECTUAL PROPERTY AND INTERNATIONAL TRADE COURT: TRADEMARK AND DOMAIN NAME CASE STUDY

In order to fulfill its obligation under the Agreement on TRIPs and to promote the strong enforcement of intellectual property disputes in international trade, the Central Intellectual Property and International Trade Court (CIPITC) is established. The proposal for appointment of a committee for setting up of pattern of the CIPITC was submitted to the Cabinet on March 22, 1993. By approval of the Cabinet on May 4, 1993, the Committee was composed of the Permanent Secretary for Justice as vice chairman, and 22 other members from public and private sectors. The Committee finished the draft Bill for the Establishment of and Procedure for the Intellectual Property and International Trade Court within one year after its appointment. The Bill was approved by the Cabinet on May 12, 1994. Subsequently, it was approved by the House of Representatives and the Senate on May 15 and August 16, 1996 respectively. The Act was proclaimed in the Government Gazette on October 25, 1996. Besides the

Act, the Committee was also the architect for the first draft of the Rules for Intellectual Property and International Trade Cases concerning proceeding and taking of evidence in court [15].

The CIPITC has power to adjudicate both civil and criminal cases regarding intellectual property and civil cases regarding international trade. Criminal cases concerning intellectual property legislation are as follows:- Offences against trademark, copyright, and patent infringement under the Trademark (including Domain name), the Copyright and the Patent Acts [16]. Although the Court has only been in existence since 1 December 1997, it has already dealt with a considerable number of cases as shown in Table 1 [17].

Table 1 Case Statistic of the Central Intellectual Property and International Trade Court: December 1, 2000 – September 30, 2003.

Case Type	2000	2001	2002	2003
International Trade	771	520	370	264
Intellectual Property (Civil Case)	102	138	157	142
Intellectual Property (Criminal Case)	2141	3252	3582	2803

Next, an example of trademark and domain name case will be discussed. Generally, the trademark owner is entitled to prevent other parties from using the trademark in respect of goods or services are of the same origin or are affiliated to or associated with or sponsored by such trademark owner. The trademark owner is accordingly entitled to prevent even an owner of noncompetitive goods or services from using the owner's trademark.

Case study: Judgment of the Central Intellectual Property and International Trade Court [18]

Case no. (Black) IP 132/2544

Case no. (Red) IP 121/2545

Intel Corporation (Plaintiff)

Intelcard Industries Co., Ltd. (Defendant)

Where the cause of this case arises out of the dispute regarding trademarks (and domain name) within Thailand, the case falls under the jurisdiction of the CIPITC, which is deemed a court competent jurisdiction. In this case, both the plaintiff and the defendant carry on the same computer business with certain overlapping objectives. The defendant uses the word "INTELCARD" to indicate as the trademark. Since the defendant's business relates to computer and is associated with or has some connection with the plaintiff's activity, the use by the defendant of the plaintiff's term "INTEL" is held to make the public think that the defendant's business is a part of or relation to the plaintiff's activity. Moreover, both the

plaintiff and the defendant keep trading via the Internet, the customers may be confused by similar trademark. Consequently, the plaintiff is entitled to prevent the defendant from using the word "INTEL" either in Thai or English version.

For the domain name aspect, the defendant holds "Intelcardgroup.com" for a domain name of the company. As earlier ruled and mentioned, the defendant has no right to use the word "INTEL" either in Thai or English version, the plaintiff is entitled to prevent the defendant from registering its company name as a domain name.

This case shows that the enforcement of trademark and domain name Rights in Thailand has been successful in a satisfied level. The newest challenge to court roles comes from the rise of Information Technology. However, not many cases in Internet issues have been brought to the Central Intellectual Property and International Trade Court in Thailand, but these additional issues will become one of the most controversial Intellectual Property issues in the near future.

7. CONCLUSIONS

The explosive growth of the Internet has resulted in a dramatic increase in the importance of domain names in the conduct of business. Businesses have traditionally developed brand name recognition of their products and services by the use of trademarks and service marks. The IT society needs to consider how best to address the conflicts between trademark rights and domain names recognizing the complexity of the range of situations that may apply and recognizing that any measures in domain name policies need to be practical and effective. This paper shows that the amended Trademark Act, the Trademarks Act (No.2) of B.E. 2543 (2000), is a very positive step forward for Thailand, setting new standards for the e-business sector to protect domain name and trademark rights.

In conclusion, Intellectual Property law is an exciting practice for Internet users and exciting area for academics. People in IT society should pay more attention on the Intellectual Property Rights to gain more understanding and to provide adequate protection.

REFERENCES

- [1] Milton Mueller "Trademarks and Domain Names: Property Rights and Institutional Evolution in Cyberspace", Syracuse University School of Information Studies, 1999.
- [2] Mark Radcliffe, "Advanced Domain Name and Trademark Issues", Jan. 2002.
- [3] Linda Rosencrance, "Bank of America Buys Loans.com Domain for US\$3M", Computerworld Online, Feb., 2000.

- [4] Gail Robinson, "You Paid How Much for That Domain Name?", *Domain Notes*, May 2000.
- [5] David Halperin, "Corporate Trademarks and the Future of Domain Disputes", *TechNewsWorld*, Nov. 11, 2003.
- [6] Russin and Vecchi, "Legal problems of domain name registration in Thailand", *International legal counselors, Thailand*, 2001.
- [7] Beth Ellyn Rosenthal, *Family Thais Lead to Bangkok.com*, Everest Partners, L.P. Report, 2000.
- [8] *Nation Multimedia Bets on Internet With Thailand.Com Site*, Bloomberg, January 24, 2000.
- [9] Kate Crabtree, "Domain Names Under ICANN: Technical Management or Policy Chokepoint", *Computers, Freedom and Privacy (CFP) Document* 2000.
- [10] Santi Rattanasuwan, "Domain Name Protection", *The Central Intellectual Property and International Trade Law Forum, Special Issue* 2000.
- [11] *An Overview of Domain Names: Domain Name Violation: Cyber Squatters & Typo Squatters*, 2003.
- [12] Michael B. Sapherstein and Staff Writer, "Cyber-Squatter Barred from Using Trademark as Domain Name", *B.C. Intell. Prop. & Tech. F.* 112104, 1996.
- [13] Wutipong Vechayanon, "Intellectual Property Rights Enforcement and the Role of Intellectual Property Court in Thailand", *The Intellectual Property and International Trade Law Forum, Special Issue* 1998.
- [14] Satyapon Sachdecha and James Pate, "Commentary: Thailand's New Trademarks Act", *Thailand International Law forum*, 2000.
- [15] *The History of the Central Intellectual Property and International Trade Court*, *The Central Intellectual Property and International Trade Court Document*, 1998.
- [16] Akarawit Sumawong, "Infringement of Intellectual Property Rights, Some Case Studies in the Field of Copyright an Related Rights: Thailand's Experience", *The Central Intellectual Property and International Trade Law Forum, Special Issue* 1999.
- [17] *Case Statistic of the Central Intellectual Property and International Trade Court*, *The Intellectual Property and International Trade Law Forum, Special Issue* 2003.
- [18] *Judgment of the Central Intellectual Property and International Trade Court: Case no. (Black) IP 132/2544 and Case no. (Red) IP 121/2545*, *The Central Intellectual Property and International Trade Law Forum, Special Issue* 2003.

Are E-Bank Related Business Methods ^[1] Patent Subject Matter under Chinese Law?

Yongjun Jin

School of Economics and Management, Tsinghua University, Beijing 100084, China
jinyj@em.tsinghua.edu.cn

ABSTRACT

In the Citibank-Patent-Case, the underlying issue is whether E-bank related business methods are subject matter of patent under Chinese law. At first flash, there are at least two barriers to patent protection for Citibank to cover. The first, E-bank related business methods only function with the help of computer software, so the coming issue is whether computer software is patent subject matter. The answer is yes, the reasons for that are.... The second, E-bank related business methods are computerized ones, so the remaining issue is whether business methods are patent subject matter. The answer is also yes, the reasons for that are.... So the position of Chinese legislature and government regulation is....

Keywords: business methods, computer programs, patent subject matter, barrier to patent, reason methods and rule

1. INTRODUCTION

On the 5th day of September 2002, a piece of news review in Southern Weekend, headlined "Citibank [2] Has Been Stealing to Apply Patent, When Will Chinese State Owned Banks Wake up from Daydream?" disclosed that Citibank had conducted 19 cases of patent application in china, and all of them involved business methods. [3] When the newspaper scattered all over the country on that day, that piece of review therein had immediately catch people' eyes, no matter who were general readers or employees in the related industries. Even the title itself had meant something. [4]

From the legal point of view, the related patents in the story are patent to invention under The Chinese Patent Act ("PA"); the related business methods, in more specific words, are computer software based business methods; nearly all patents involve E-bank industry, so they may be called E-bank related business method patent. [5]

The issue underlying that story is whether E-bank related business methods are subject matter of patent under Chinese law. [6] This note will discuss this issue. [7] For discussion purposes, Part II will describe the conception of invention under patent law in simple words; Part III will focus on barriers to patent, mainly the barrier of reason rules and methods; And in this part business methods and computer software or programs themselves will also be discussed. Part IV analyzes in two perspectives whether E-bank related business methods are patent subject matter under Chinese law. The final part will give the conclusion to this note.

2. INTIOION UNDER THE CHINESE PA

2.1 Invention

Under the Chinese PA, invention is one of the three kinds patent subject matter. [8] But what is an invention?

In simple words, invention is a kind of technical solution. [9] For example, the traditional Chinese paintings are mostly drawn on one kind of paper called "Xuanzhi". Because of the special property of that kind paper, it is very difficult to print these paintings in Xuanzhi with modern print machine. This is a time-consuming and unsolved technical puzzle. There is a hero, who is a Shenzhennese, called Zhuang Zhihe, and he has succeeded in achieving a solution. With distinctive arrangement of printing process, the similar machine the industry has employed can print the traditional Chinese paintings in Xuanzhi. [10] What Zhuang Zhihe has achieved is a technical solution for a concrete technical issue. Invention under patent law is such kind of solution. [11]

It is necessary to remind, the cited invention above involves a solution to process of product. An invention may also be a solution concerning products themselves. The exact example is the solution concerning the traditional Chinese paintings [12] printed with the patented process. [13]

If we say those two kinds above are groundbreaking ones, then an invention under patent law may be an amendment to either of them too. For example, an amendment to the printing process or printed product for which Zhuang Zhihe claimed a patent. [14]

It is with those three kinds of invention in mind that the Chinese Implementation Rule of Patent Act("IPA") article 2 section 2, using legal terminology, generalize like this : "Invention is a ... technical solution concerning product, process or amendment to either of them."

2.2 Disclosing files to invention

In order to apply a patent for an invention, the applicant shall describe the invention as required and disclose. The required writings include abstract, specification, claims and so on. And the claims is the most important

one of them, since it is this file that describes the technical characteristics of the related invention. [15] For example, with regard to the process invention of Zhuang Zhihe, the claims to that states in such a structure: "a process for plate making and printing... the characteristics of it is..." [16]

The claims structure of products invention is similar. For example, the Chinese Patent No.90101931 is a process and product combined patent. Claim 1 to that Patent claims a solution about a battery and describes in the similar structure like this: "a battery is made up of... the characteristics of it is..." [17]

In a word, it is through describing the characteristics of the involved invention that the claimed invention is circled with bright lines from the prior art.

3. BARRIERS TO PATENT PROTECTION

Under the Chinese PA there are active and passive requirements for patent protection. The active requirements are novelty prong, non-obviousness prong and usefulness prong; the passive requirements are barriers or exceptions to patent protection. [18] What this note will involve is one kind of barrier, i.e., the barrier of reason rules and methods.

3.1 Barrier of reason rules and methods

Reason rules and methods are rules and methods for thinking, recognizing, judging or remembering of human being. [19] In any sense they are not technical solutions, so they are not invention in the sense of patent law. [20] Just because of that, the Chinese PA holds that they are barrier to patent protection. [21] And those rules and methods include but are not limited to: special methods for patent examination, methods and systems for organizing, producing, executing or managing, computer languages or algorithms, statistic, accounting or recording methods, computer programs and so on. [22]

Another step further, the Examination Guide clarifies, if an application claims reason rules and methods only, the applied patent cannot be issued; however, an application claims something else besides reason rules and methods, and the decisive factors in that application are not the said reason rules and methods, the patent can be issued, that is, that application claims the patent subject matter. [23]

3.2 Computer programs

One item of reason rules and methods is computer programs. [24] But what is a computer program? It will be pretty well that we take system program as an example to explain this issue. A system program is mainly a unit of integrated three parts. First, is User Interface; second, is hardware and software drive; third, is data process. And the third is the most important one. After a user has driven the hardware and software, he

will enter some data into the computer via the User Interface, and these data be processed by the data process. This process is made up of two parts, first is data structures and second is algorithms. The data structures will organize all input data for the algorithms to process and the algorithms will process all input data organized by the data structure first. [25]

The data process is written with computer languages. The languages may be assembly languages, high-level languages or something else. A program in one language of them is a program in source code. The basic unit of the program is statement. After the program is entered into a computer, it will be translated into machine languages, and the result from that translation is a program in object code. The basic unit is instruction. [26]

So the Chinese Regulation of Software Protection article 3 subsection 1 provides: "Computer program, refers to a sequence of coded instructions to achieving a specific result which can be executed by such devices as computers that have information-processing capacity, or a sequence of symbolized instructions or symbolized statements which can be automatically converted into a sequence of coded instructions. The source code text and object code text of the same program shall be deemed to be the same and one work." Examination Guide adopts this definition with the only one change that it substitutes the second sentence in the definition with this one: "Computer program includes source code and object code." [27]

With regard to whether a computer program is subject matter of patent, although Examination Guide holds that computer program is a special item of reason rules and methods, but separates it from the other items and deals with independently in other chapter. [28]

3.3 Business method

Maybe the conception of business method was originally from the U.S. federal cases of patent. In patent cases judges sometimes used "business methods" and sometimes "methods of doing business", but they provided no clear definition on either of them. [29] The earlier U.S. federal cases had involved hotel security, [30] method for parking cars, [31] and bank account. [32] As far as these cases are concerned, some items of reason rules and methods under Chinese law at least are equivalents to the U.S. business methods. For example, methods and systems for organizing, producing, executing or managing, statistic, accounting or recording methods and so on. [33]

In the U.S., before *State Street Bank & Trust Co. v. Signature Financial Group, Inc.*, [34] it was only some judges who accepted barrier of business methods to patent protection only in dicta. But conventional wisdom and hornbook law had hold that "methods of

doing business" are not patentable, [35] they were exception to patent subject matter.

4. E-BANK RELATED BUSINESS METHODS

With the appearance and development of the Internet, besides the traditional business methods, a new generation of business methods came into existence and became popular. This new generation was computer program based business methods. In general, these business methods may be methods for producing, distributing or accounts recording with the help of the Internet. [36] The business methods for which were applied patent by Citibank in china are mostly computer program based business methods, [37] and much more complicated ones. For example, the electronic money system for which Citibank applied patent and claimed the right of priority was a pending patent in China. Its Claim 1 described a system for transferring electronic notes between processor-based electronic modules; and Claim 5 described a method for transferring electronic notes between processor-based electronic modules. [38] Both of them were Chinese versions translated from the corresponding claims of the U.S. patent [39] upon which the related Chinese application claimed the priority right. [40] Claim 1 of the U.S. patent claimed a product invention and Claim 5 a process invention. [41] It is what Claim 5 of the U.S. patent claimed that deserves discussion. The technical solution claim 5 claimed is one hundred percent a business method. A method for transferring electronic notes between processor-based electronic modules comprises the steps of: establishing a secure session, creating a transfer electronic note and transferring the said transfer electronic note. This process does not work without either of hardware and software, for example, processor-based electronic modules are hardware; the cryptographic algorithms used to establish secure session, are software. But it is neither the hardware nor software themselves it employs. [42] In other words, the software and hardware are necessary for operation of the said process, but the technical solution Claim 5 claimed consists in the process, drives nothing from either of the software and hardware. Of course, the process itself may be integrated into a computer program, [43] but that is another issue. Such kind of E-bank related business methods is business method in its strict sense. [44] From this point of view, that Citibank applied 19 patents for its business methods is not an exact description. [45]

The decisive issue is whether such kind of E-bank related business methods is patent subject matter under Chinese law. Let alone the general requirement for patent protection, [46] from the analytic-logical point of view, there are two relied sub-issues that shall be discussed separately. First, Are computer programs patent subject matter? Second, Are business methods patent subject matter?

4.1 Are computer programs patent subject matter?

(A) Legal position of China

First, the Chinese PA considers reason rules and methods as barrier to patent protection. Examination Guide interprets that reason rules and methods include computer programs. [47] In China Examination Guide is one of government regulations issued by the Patent Office, not statutes passed by the People's Congress. Although there is such a piece of interpretations in Examination Guide, but what it expresses is not bound to be the legal position of China. To our relief, before promulgation of the original version of Examination Guide, [48] there really were decisions of the People's Court that shared the common idea. [49]

Second, Examination Guide chapter 9 covers the application of computer program based invention. [50] This chapter holds that, if an application claims computer programs only, for example, a program recorded in ROM, PROM, VCD, DVD, the related patent cannot be issued; However, an application does not claim computer programs only, and what is claimed intends to resolve technical issues, takes technical measures, and produces technical results, then it is not permitted that the related patent be refused just because of computer programs. [51]

For example, Chinese character coding methods, although they can be used in lexicography and characters searching, but they do not intend to resolve technical issues, take technical measures, or produce technical results. Even if being programmed into computer programs, they are reason rules and methods. Otherwise, with the help of Chinese character coding methods and employing keys board of computer of general purpose, a method for entering Chinese characters into computer can be achieved. This method performs the function of characters inputting and something else. Such a design involves computer programs but itself is a solution for a concrete technical issue. So it is patent subject matter, or patentable. [52]

A little regret that, all above restated are only supported by the administrative regulations. But till now none of judicial decisions says no to that.

(B) Reasonableness of that position

Firstly, some scholar held such a position was also accepted by some other Patent Offices in this world, especially, in developed countries. [53] For example, in the U.S., the counterpart exception to patent protection for computer programs is mathematical algorithms [54] and computer programs are mainly algorithms or aggregate of algorithms. [55] As far as this doctrine is concerned, there is no special provision in the U.S. PA, but the U.S. Supreme Court had heard three related cases, and they were *Gottschalk v. Benson*, [56] *Parker v. Flook* [57] and *Diamond v. Diehr*. [58] Although the

Supreme Court did not interpret that doctrine extremely in similar language at these three cases, but the U.S. Court Appeals for the Federal Circuit ("CAPC") at *State Street Bank & Trust Co. v. Signature Financial Group, Inc.* [59] had expressed very clear in 1998 that, mathematical algorithms are not patentable to extent that they are merely abstract ideas, [60] however, a practical application of mathematical algorithms which produces a useful, concrete and tangible result is patentable. [61] Since the Supreme Court denied certiorari on January 11, 1999, that court's opinion had been the law of the land. [62] So the Chinese and U.S. legal positions are different in approach but equally satisfactory in result.

Secondly, the decisive reasonableness consists in computer programs themselves. We may take Adobe Acrobat 5.0 as an example. An end user uploads this program in an intranet, and users in that intranet can download it freely. Such a way of using programs, just involves one hand, using programs as literary work. [63] And another end user downloads and installs it in his computer with Windows XP, he operates the program and transfers a word file called Word-file into a PDF file called PDF-file. Although all steps work automatically, [64] but automate cannot change such a fact that from one product (Word-file) he produces another product (PDF-file). This does mean the program is a process of product. Such a way of using programs, involves another hand, using programs as process of product. In a word, the program has two faces, one face tells us literary work; another face tells us process of product. [65] From the face of literary work, copyright protection shall be accessible for that program; from the face of product process, patent law shall protect that program. [66]

Thirdly, as literary work, computer programs enjoy copyright protection. In nature, what copyright law protects is the expression of computer programs in language, such as machine languages, assembly languages, high-level languages or something else. As process of product, computer programs enjoy patent protection. What patent law protects is the technical solution that consists in computer programs, performing some function or processing something. Computer programs are expressed in languages, but designed to perform some technical function or process something. So enjoying both of them is not conflicting, on the contrary, necessary. [67]

In a word, computer programs as process of product are patent subject matter in China.

4.2 Are business methods patent subject matter?

(A) Legal position of China

In China there is no independent business methods exception. Some items of reason rules and methods are

counterpart to business methods. [68] But there are no judicial decisions to make sure whether business methods may be listed adversely as items of reason rules and methods. For discussion purposes, here presumed they may.

As far as reason rules and methods are concerned, in much more definite words, Examination Guide does hold that, if an application claims reason rules and methods only, the applied patent cannot be issued; but an application involves reason rules and methods, the solution described in the application consists in something other than reason rules and methods themselves, the applied patent can be issued. [69] Such a position is similar to the one concerning computer programs. [70] In a word, if business methods intend to resolve a technical issue, take technical measures and produce technical results, they are patent subject matter. Since conception of business methods is from the U.S. origin, before going further we would restate the legal position of the U.S. first.

(B) Legal position of the U.S.

As far as barrier of business methods is concerned, the U.S. PA is a sister to the Chinese PA. It does not utter a word about that doctrine. The U.S. common law also refuses to hold business methods being exception to patent protection. In *State Street Bank & Trust Co. v. Signature Financial Group, Inc.*, [71] CAPC put an end to that doctrine absolutely which had not really existed in the legal world.

Firstly, CAPC held that, since its inception, business methods exception had merely represented some general, but no longer applicable legal principle, perhaps arising out of the requirement for invention. [72] Since 1952, business methods had been subject to the same legal requirement for subject matter applied to any other process or method. [73] It was time to let business methods exception rest forever. [74]

Secondly, CAPC insisted that, business methods exception had never been invoked by CAPC or the former of that court initialed "CCPA". Even if the issue of subject matter was concerned, the court invalidated patents on Abstract Ideas exception [75] created by The Supreme Court. [76]

Lastly, CAPC held: "Even the case frequently cited as establishing the business method exception to statutory subject matter, *Hotel Security Checking Co. v. Lorraine Co.*, 160 F. 467 (2d Cir. 1908), did not rely on the exception to strike the patent." [77] Instead, the patent was declared invalid for lack of novelty and invention. [78]

(C) What is the reason for that position?

Barriers to patent protection in patent law can be

classified into two groups, first, patent subject matter is invention, so what is not invention is barrier to patent protection; second, even if what is claimed in a patent application is an invention, in consideration of legislation policies such as national security patent act forbids to issue patent for that invention. Then all that are forbidden are also barriers. The former, for example, the scientific discovery listed in the Chinese PA Article 25; the later, for example, a matter gained by methods of atomic nucleus that is also listed in the Chinese PA Article 25. [79] Business methods, if they are, are the former. [80]

In the examination proceedings for patent, the Chinese PA Article 25 that lists barriers to patent protection is examination item to the preliminary examination or substantial examination. And the Chinese PA Article 22 that provides novelty prong, non-obviousness prong and usefulness prong is examination item to the substantial examination. [81] For any patent application, the preliminary examination is logically earlier than the substantial examination in time point. So the first step is make sure if an application claims invention; if yes, then the second step is make sure if what is claimed may pass the novelty prong, non-obviousness prong and usefulness prong.

From the two perspectives above, it is sufficient to conclude that, the issue whether business methods are barrier substantially is the issue whether business methods are invention. As far as the later issue is concerned, the conclusion in China and the U.S. are common. If they are technical solution, they are invention under patent law. [82] Logically, it is a sound conclusion. But the decisive word in that statement is "if". If no if, how about that? Are business methods absolutely not invention? Or business methods may be invention? This is a factual issue that needs evidence to confirm.

(D) Is there evidence for that conclusion?

Take the above examples first to prove that business methods may be invention. We have discussed the printing method of Zhuang Zhihe that has been issued Chinese patent. That printing process is a traditional product process. [83] And we also have discussed Adobe Acrobat 5.0 that has received the U.S. patent, a business method to produce one product (PDF-file) from another (Word-file). [84] Let alone the later was patented in the U.S., as far as these two examples are concerned, there is not so much hesitation for us to say that they are members of one family, business methods may be invention.

Perhaps the only pending doubt is that Adobe Acrobat 5.0 is software for office purpose, not the traditional product process. Even if definitely so, this factor is not decisive. Because the patent subject matter is invention, and invention is technical solution. May software only

for office purpose be technical solution? Of course, yes. Then that software may be invention in patent law. In fact, there is another kind of software for office purpose, a method for entering Chinese characters into computer, and that is patented in China. [85]

And the U.S. Patent Office has issued patent for business methods for a long time. In 1779, first business method patent was issued to Jacob Perkins of Massachusetts for "detecting counterfeit notes". Because of a fire in the Patent Office in 1836, all materials about that patent were lost. The first business method patent all files of which were kept well today was patent issued to John Kneass in 1815 titled "a mode of preventing counterfeiting". And from then on, a lot of patents have been issued for business methods. [86] All these are indirect evidence for the conclusion that business methods may be invention.

All in all, business methods, if intend to resolve technical issues, take technical measures and produce technical results, they are patent subject matter.

5. CONCLUSION

In the Citibank-Patent-Case, the basic issue is whether E-bank related business methods are patent subject matter under Chinese law. The latest Chinese law and regulation hold that, computer programs upon which performance of E-bank related business methods depends, may be technical solution, then they are patent subject matter; and business methods, although are computerized ones in E-bank industry, but may be technical solution, then they are patent subject matter too. So E-bank related business methods are patent subject matter. Although China is a developing country, but in this special field of business method patent, China shares common property ideas with developed countries, for example, the U.S.

With regard to whether a patent may be issued for an individual business method, how that patent is issued, and whether there are some specially applied restrictions for that patent, all these topics afford another note. [87] (10/23/03)

FOOTNOTES

[1] See infra text accompanying note 3, infra note 5 and accompanying text, and infra text accompanying notes 41-44.

[2] The original Chinese words in that piece of news review are "Citibank", but the real applicant to patents is Citigroup. This note will keep on using the word "Citibank". See Wang Feng and Ma Ling, Citibank Has Been Stealing to Apply Patent, When Will Chinese State Owned Banks Wake up from Daydream?, Southern Weekend, September 5, 2002, <http://finance.sina.com.cn/g/20020905/1444251033.html> (visited Oct. 7, 2003).

- [3]Id.
- [4]See e.g., Han Fenglei, Citibank Entering with Advantage of Patent, The Central Bank Calling for Big Four to Resist, Chinese Management Weekly, December 19, 2002, <http://finance.sina.com.cn/b/20021219/1717292233.shtml>(visited Oct. 7, 2003).
- [5]The author has used "citi" as key word to applicant field, searched in the Chinese patent database (<http://www.sipo.gov.cn>), and checked the search results one by one. The final conclusion is that.
- [6]See Wang and Ma, supra note 2.
- [7]But the decisive word in that statement is "whether". This note will only focus on this point. If yes, and in consideration of legislation policies, whether the legislator can establish a barrier to patent is another issue. This note will not discuss this issue here. See generally infra text accompanying note 79.
- [8]See Patent Act [PA] art. 2(China).
- [9]See Implementation Rule of Patent Act [IPA] art. 2 section 2 (China) ("Invention is a ... technical solution concerning product, process or amendment to either of them.").
- [10]<http://www.sipo.gov.cn/sipo/zljs/default.htm> (visited Oct. 8, 2003).
- [11]In fact, his method is a patented one. See the Chinese Patent No. 93114279 (issued Dec. 30, 1995).
- [12]See the Chinese Patent No. 93114279 (issued Dec. 30, 1995).
- [13]See supra note 11.
- [14]See supra notes 9-13 and accompanying text.
- [15]See IPA art. 20 section 2(China)("The claims shall describe the characteristics of invention...so that the protection extension is simply and clearly lined.").
- [16]<http://www.sipo.gov.cn/sipo/zljs/default.htm> (visited Oct. 8, 2003).
- [17]<http://www.sipo.gov.cn/sipo/zljs/default.htm> (visited Oct. 8, 2003).
- [18]See generally Jin Yongjun, The Principle of Intellectual Property Law 33-51(2002).
- [19]See Examination Guide Part Substantial methods"[hereinafter Examination Guide 2.1.3.2] (china)(being promulgated by the original Patent Office in 1993).
- [20]Id.
- [21]See PA art. 25 section 1 subsection 2(china)(reason rules and methods not being issued patent).
- [22]See Examination Guide 2.1.3.2(china).
- [23]Id.
- [24]See supra text accompanying note 22.
- [25]See Jin, supra note 18, at 270-71.
- [26]See Jin, supra note 18, at 271-72.
- [27]See Examination Guide 2.9.1(china).
- [28]See Examination Guide 2.1.3.2(china). See also infra Part IV.1.A.
- [29]See, e.g., William D. Wiese, Death of a Myth: the Patenting of Interest Business Models after State Bank, 4 Marq. Intell. Prop. L. Rev. 17, 31 &n.86 (2000).
- [30]Hotel Security Checking v. Lorraine, 160 F. 467 (2d Cir. 1908).
- [31]Loew's Drive-In Theatres Inc. v. Park-In Theatres, Inc., 174 F.2d 547, 81 U.S.P.Q. (BNA) 149 (1st. Cir. 1949).
- [32]In re Wiechers, 347 F.2d 608, 146 U.S.P.Q. (BNA) 52 (C.C.P.A. 1965).
- [33]See supra text accompanying note 22.
- [34]149 F.3d 1368, 47 U.S.P.Q.2d (BNA) 1596 (Fed. Cir. 1998).
- [35]See Wiese, supra note 29, at 30-31.
- [36]See Michael L. Fuelling, Manufacturing, Selling, and Accounting: Patenting BMs, 76 J. Pat. & Trademark Off. Soc'y 471, 473 (1994).
- [37]See supra text accompanying note 5.
- [38]<http://www.sipo.gov.cn/sipo/zljs/default.htm> (visited Oct. 8, 2003) .
- [39]In the U.S. Patent No. 5,799,087 (issued Au. 25, 1998), the applicant claimed: "1. A system for transferring electronic notes between electronic modules comprising: electronic modules each having a processor, a memory, and the capability to create a cryptographically secure channel and transfer and receive electronic notes via said cryptographically secure channel, and where each electronic module stores said electronic notes in its respective said memory; wherein each stored electronic note comprises: a body group of data fields including data indicative of a monetary value associated with said electronic note; a transfer group of data fields including a list of transfer records, where each transfer record is generated by a transferor electronic module and includes a sequence number that distinguishes a transferred electronic note from one or more other transferred electronic notes transferred from a common said transferor electronic module and generated from a common electronic note." "5. A method for transferring electronic notes between processor-based electronic modules, comprising the steps of: establishing a cryptographically secure session between a transferor electronic module and a transferee electronic module; creating a transfer electronic note by appending a transfer record to an electronic note stored in said transferor electronic module, where said transfer record indicates whether all or some portion of a monetary amount of said electronic note is being transferred; transferring said transfer electronic note from said transferor electronic module to said transferee electronic module; and wherein said transfer record includes a sequence number that distinguishes said transfer electronic note from another transfer electronic note from said transferor electronic module." <http://www.uspto.gov/patft/index.html> (visited May 5, 2003) .
- [40]<http://www.sipo.gov.cn/sipo/zljs/default.htm> (visited Oct. 8, 2003).
- [41]See supra note 39 and supra text accompanying notes 15-17.
- [42]See supra note 39.
- [43]See, e.g., infra text accompanying notes 62-65.

- [44]See David T. Dutcher, *Patents on Methods of Doing Business*, 79 *Denv. U.L. Rev.* 173,187 (2001).
- [45]See *supra* text accompanying note 3.
- [46]See *supra* text accompanying note 18.
- [47]See *supra* Part III.1.
- [48]The promulgating year was 1993. The outstanding version was done in 2001.
- [49]See Chen Yongshun et al, *The Selective Cases of Patent Administrative Litigation(1988-1998)* 1-20 (1999).
- [50]Computer programs based invention, refers to, whole or partly, computer program based solution for technical issues. See Examination Guide 2.9.1.
- [51]See Examination Guide 2.9.2.
- [52]See Examination Guide 2.9.3. And see, e.g., the Chinese Patent No. 97120110.
- [53]See Pu Manwen, *The Protection of Computer Programs Involved Invention*, *IP Studies* 49(1998)(in Chinese).
- [54]See *infra* text accompanying note 60.
- [55]See *supra* text accompanying note 25.
- [56]409 U.S. 63 (1972).
- [57]437 U.S. 584 (1978).
- [58]450 U.S. 175, 177 (1981).
- [59]149 F.3d 1368, 47 U.S.P.Q.2d (BNA) 1596 (Fed. Cir. 1998).
- [60]Abstract ideas are one kind of exception to patent protection created by the Supreme Court. See *infra* note 75.
- [61]See *State Street Bank & Trust Co. v. Signature Financial Group, Inc.*, 149 F.3d 1368, 1373(Fed. Cir. 1998).
- [62]See Christopher S. Cantzler, *State Street: Leading the Way to Consistency for Patentability of Computer Software*, 71 *U. Colo. L. Rev.* 423(2000).
- [63]See *supra* Part III.2.
- [64]See *supra* text accompanying note 27.
- [65]See generally Vincent Chiappetta, *Patentability of Computer Software Instruction as an Article of Manufacture: "Software am such as the Right Stuff"*, 17 *J. Marshall J. Computer & Info. L.* 89,141-43(1998).
- [66]In fact, Adobe Acrobat 5.0 is a patent product. The involved patent is the U.S. Patent No. 4,558,302 (issued Dec. 10, 1985) .
- [67]In fact, the U.S. Patent No. 4,558,302 (issued Dec. 10, 1985) , had not claimed source codes or object codes, but the process that consisted in the program. For example, the claim 107 says : "In a data compression and data decompression method, a compression method for compressing a stream of data character signals into a compressed stream of code signals, said compression method comprising the steps of storing, in the locations of a memory, strings of data character signals; searching said stream of data character signals; inserting into said memory; assigning a code signal corresponding to said stored extended string, and providing the code signal associated with said longest match so as to provide said compressed stream of code signals." <http://www.uspto.gov/patft/index.html> (visited May 5, 2003) .
- [68]See *supra* Part III.3.
- [69]See *supra* Part III.1.
- [70]See *supra* Part IV.1.A.
- [71]149 F.3d 1368, 47 U.S.P.Q.2d (BNA) 1596 (Fed. Cir. 1998).
- [72]Before the Amendment made to PA in 1952, there was no non-obviousness prong in the U.S. PA although that prong existed now. But in common law, there was requirement for invention although it was not so clearly defined. The Amendment to PA 1952 substituted the non-obviousness prong to requirement for invention. See Donald S. Chisum et al., *Principles of Patent Law* 515-16(2001).
- [73]See *supra* note 72.
- [74]See *State Street Bank & Trust Co. v. Signature Financial Group, Inc.*, 149 F.3d 1368, 1376(Fed. Cir. 1998).
- [75]The Supreme Court had identified three categories of subject matter that are not patentable, namely laws of nature, natural phenomena, and abstract ideas. See *Diamond v. Diehr*, 450 U.S. 175,185 (1981).
- [76]See *State Street Bank & Trust Co. v. Signature Financial Group, Inc.*, 149 F.3d 1368, 1376 (Fed. Cir. 1998).
- [77]See *id.* at 1377.
- [78]See *id.* at 1377.
- [79]See Jin, *supra* note 18, at 45-50.
- [80]Logically, they may be the later, but that means nothing to the statement this note focuses on. See *supra* note 7.
- [81]See Jin, *supra* note 18, at 148-49.
- [82]See *supra* Part IV.2.A-B.
- [83]See *supra* Part II .1 and *supra* note 11.
- [84]See *supra* text accompanying notes 62-65 and *supra* note 66.
- [85]See *supra* note 52 and accompanying text.
- [86]See the U.S. Patent and Trademark Office, *White Paper on Automated Financial or Management Data Processing Methods (BMs)*, <http://www.uspto.gov/web/menu/busmethp> (visited May 26, 2003).
- [87]See, e.g., Dutcher, *supra* note 44, at 186, 189-92 and 196-99.

Continuous Audit Agent System

Charles, L. Y. Chou, Timon, C. Du

Department of Decision Sciences and Managerial Economics, the Chinese University of Hong Kong,
Hong Kong, China
{lychou, timon}@baf.msmail.cuhk.edu.hk

ABSTRACT

Recently, the provision of the real-time accounting reports over the Internet for the public corporation became popular. In response to this change on the service requirement, audit professionals considered to provide "continuous auditing" to diminish or minimize the lag time of the provision of an audit opinion behind the occurrence of business events. Though has been proposed the concept at 80's, the provided implementation models of continuous auditing are still far beyond feasible. Major obstacles to hinder wide adoption in industry lie on two issues. (1) The development of continuous auditing system should be independent to the development of client's information system. (2) The audit procedures that could be implemented by agent base automatic audit system should not be limited to the analysis on corporation internal data. In this paper, an agent base continuous auditing model and system is proposed to tackle these two major concerns. Several kinds of mobile agents are initiated by an audit agency system to simulate the behaviors of a human auditor on the audit procedures to collect various audit evidences. Each kind of mobile agent simulates a specific kind of audit procedure and acts on behalf of a human auditor to access and to inspect audit-related information hosted in the distributed information sources

Keywords: continuous auditing, multiple agent base system

1. INTRODUCTION

In recent decades, the adoption of information technology and electronic business model has dramatically changed the way we manage business and business information. Consequently, audit professionals are confronting pressing pressure to retrospect the way they have been doing auditing for decades. The pressure is coming from two directions: (1) While implementing electronic business model, the documentation for the business communication among participating parties and the information to support business operation within corporation are generated and collected in electronic forms. Under such circumstances, more and more audit evidence, which is taken by audit professionals as verification support to validate the corporation statement in their auditing process, could be collected in electronic form. (2) It is getting more corporations to disseminate their financial reports over Internet in real-time basis. The improvement in information efficiency has provided decision makers convenience in collecting useful information online. Nevertheless, such improvement turns out to be the pressure to the audit professionals to provide real time auditing service (Robert K. Elliott, 2002). In responding to this demanding request, "continuous auditing" was therefore proposed as an alternative solution. The continuous auditing here represents the assurance service where the time between the occurrence of events underlying the particular subject matter of a client and the provision of an auditor's opinion on the fairness of the disclosure of the information for the subject matter is diminished, or is minimized to an acceptable level (Alexander Kogan et al., 1999, Zabihollah Rezaee et al., 2002, Zabihollah Rezaee et al.,

2001).

The research for the continuous auditing models has been on-going for two decades (Roger Debreceeny et al., 2003, S. Michael Groomer and Uday S. Murthy, 1989, Harvey S. Koch, 1981, Miklos A. Vasarhelyi and Fern B. Halper, 1991). The proposals for implementing continuous auditing are executed mostly by monitoring the generation of transaction data or by analyzing the transaction data with a set of rules defined for auditing purpose. Violations to the rules are found within the transaction lifecycle or shortly after the completion of the transaction. Then certain alarms or "red flag" should be raised to call the auditor's attention in corresponding to the exceptional situation (Yining Chen, 2004). Based on the collected evidence, Auditors not only could pinpoint the weakness spot of internal controls, but also provide improving method to the internal controls of the operation systems. However, the proposals for installing facilities for continuous auditing limit their actual implementation. Commonly proposed methods for implementing continuous auditing could be classified in two approaches: embedded audit modules (EAM), and audit data mart (ADM). Both implementations lead to certain limitations for wide adoption in industry. The limitation for EAM is at its installation. Additionally, Auditors has problems to implement EAM models independently and has difficulty to embed the modules into clients' system after system being built up. The implementation of ADM is its constraints on the abilities on collecting various types of audit evidence.

In this study, a new continuous auditing model, which uses multiple mobile agents to simulate the behaviors of a human auditor on the audit procedures to collect

various audit evidences, is proposed. Each kind of mobile agent simulates a specific kind of audit procedure and acts on behalf of a humane auditor to access and inspect audit-related information hosted in the distributed information sources. As a whole, the numerous agents behave aggregately as an audit agency system. In addition to the benefits generally expected for the intelligent agent base system (Danny B. Lange and Mitsuru Oshima, 1998), we argue that this audit agency system would overcome the limitation and the concern of EAM and audit data mart. The obvious benefits provided by proposed system are listed as below:

(1) Audit procedures can be implemented continuously—The mobile agent simulates an inspector on the audit-related information, and not embedded as a sub-routine of the information system. Therefore, the mobile agent is not activated automatically by the events on the on-going basis. Rather, the mobile agent is scheduled to execute predefined audit procedures autonomously and periodically. Through a periodical inspection on the audit-related information, the audit agency system provides timely information, e.g. daily or weekly reports, depending to actual implementation requirement for various kinds of audit evidence.

(2) The development of the audit agency system is independent to the development of the information system – As for the mobile agent is a dispatched software object to the client's information system, instead of being embedded in the information system, the audit professionals could develop the audit agent system without the intensively involving the client's legacy systems. That is, to minimize the interference to the legacy system, only the authorized secure interface to the read-only access privilege to the audit related information is needed.

(3) The software agent is flexible to access various kinds of information, and the audit agency system is easy to adapt to the change of audit environment – A dispatched mobile agent interacts with local agents to acquire information. The behavior of the local agent is instructed by the local schema the needs of the mobile agents. Note that the local agent is also a software objects in which it is programmed, generated and resided in the local system. The design of the local agent provides some advantages. For example, when the audit environment changes, e.g. the changes of audit criteria, operation system, and information system, the local agent can be easily updated individually to adapt to the new environment.

(4) The mobile agents could be bestowed with different degree of intelligent activities corresponding to the needs of implementing the audit procedures. Therefore, the ability for mobile agents in executing audit procedures is not limited to the inspection of historical record in accounting information system. Rather, they

can do more advanced activities such as comparing the information consistency in various applications, carrying information to the third parties system for confirmation, consulting inventory database for most updated stock level, etc. Apparently, the capability of auditing is thus expanded by the mobile agent programs.

2. AGENT BASE CONTINUOUS AUDIT MODEL

A software agent is a software object that is able to achieve a goal by performing actions and reacting to events with prearranged activities in the dynamic environment (Timon C. Du et al., 2004, Alfonso Fuggetta et al., 1998). Typically, software agents are programmed to have motilities and certain degree of intelligence. The mobility implies that the agents are able to transport themselves to given service platforms through the network while the intelligence encapsulates the artificial intelligence. Therefore, intelligent mobile agents are able to provide sophisticated computational or behavioral model to the remote data sources while interacting with the local services or applications. In this case, the mobile agent paradigm benefits the applications like electronic commerce, secure brokering, distributed information retrieval, parallel processing, monitoring and notification (Jon Woodroof and DeWayne Searcy, 2001), telecommunication networks services, workflow applications and groupware (Danny B. Lange and Mitsuru Oshima, 1998). Successfully developed prototypes of the applications include supply chain SMART project (<http://smart.npo.org>), brokering service in electronic market (Timon C. Du, et al., 2004), virtual enterprise (Anuj K. Jain et al., 1999), information retrieval and internet-auction house (Tuomas Sandholm and Qianbo Huai, 2000), and distributed network management (Timon C. Du et al., 2003). In the following sections, we will successively describe the major components of the multiple audit agents system for the implementation of continuous auditing.

2.1 The agent platform

In this study, we select a platform of IBM's Tokyo laboratory (<http://www.trl.ibm.com/aglets/>) for developing the audit agent system. The platform has a set of abstract classes and application programming interfaces (API) for users to build their own agent platform. The agent platform provides an environment of agent servers. Among those abstract classes, Aglet, AgletProxy, and AgletContext are the three most relevant key components to the agent activities -- creation, clone, dispatch, messaging, and disposal. Elaborating the functions of these three classes can help us to get a better picture of the agent platform (Danny B. Lange and Mitsuru Oshima, 1998).

(1) Aglet: Aglets is an abstract java object proposed by IBM development team. The name is a composite word

coming from “agent” and “applet”. It is programmed to implement all basic agent activities we expected for the mobile agents. A system developer can model the behavior of their own mobile agents by Extending this class definition.

(2) Proxy: AgletProxies are the representatives of the Aglets. In stead of directly accessing to the Aglets public methods, the intra-Aglets messaging and the communication between Aglet and the environment all go through the interfaces of AgletProxies. This design serves for two purposes: (a). it serves as a shield to protect the aglet from malicious access and (b). the proxy also provides location transparency for the aglets, that is, it can hide the real location of Aglets. Thus, The Aglet and its proxies (more than one copy is allowed) are separated across a network. Additionally, the exchanges of message with the Aglet’s proxies are just made as the interaction with local service agents.

(3) Context: Conceptually, a context is the Aglet’s workplace. Every active Aglet must be hosted by an AgletContext. The function of the AgletContext is different from a server; it is a stationary object that provides a means for maintaining and managing running Aglets in a uniform execution environment. The interface to the AgletContext is used by Aglet to get information of active Aglets and their environment factors. Basically, one computer network node can run many Aglet servers (engines), one servers can host many AgletContexts, and one AgletContext could host many Aglets.

2.2 The audit agency system

In this study, the audit agent system uses the Aglet agent platform. The system functions include taking and executing the order for audit procedures from auditors to the end of consolidating the evaluation results as the reply to the auditors. The proposed system can offload human auditors’ workload on labor-intensive works, and can benefit audit professionals a reliable and cost effective system to continuous and automatic execution of audit procedures.

Fig 1 shows the operations of the audit agent system. The main component of the system is the Agency Services Server (ASS), which interprets the command from auditors and initiates necessary mobile agents to complete the assignment. Then, various kinds of mobile agents are spawned to perform operations on different information sources, according to specialties of each type of audit evidences. Examples of collecting available audit evidences are as follows:

(1) Process definition confirmation: The mobile agent compares the current process definitions with the historical records and replies with the discrepancy. Such process definitions are maintained by the process control system, such as workflow management system

(WfMS) or enterprise resource processing (ERP), and are identical to the conventional operation. This audit procedure assures the consistency of the business operation and alerts the auditors about the changes of the operation if any. During implementation, the mobile agent carries the process definition to the clients’ site and compares against information collected from local system (agent No 1 in figure 1)

(2) Query data and precede analytical process: The mobile agent performs analytic procedures on data to check whether the inter-relationship between documents, analytical index, and company policy are

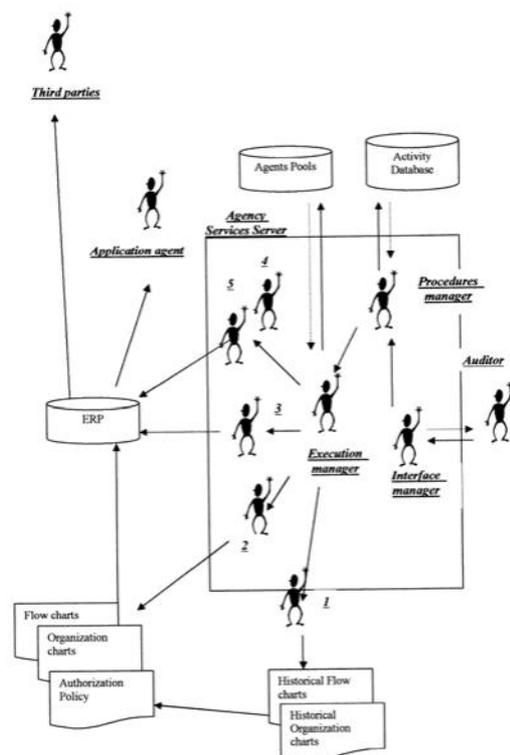


Fig 1 Audit Agency System

satisfied. By doing so, document with errors or suspicious to potential fraudulence can be allocated. During implementation, the mobile agent might visit many distributed information resources to collect related information then activate its pre-programmed analytic procedures (agent No. 2 in figure 1), or the agent can stay in ERP and complete its analytical process on the documents hosted in ERP (agent No. 3 in figure 1).

(3) Query data and take the data to third parties, e.g. customers, vendors, or bank, for further confirmation: The mobile agent retrieves and carries relevant information to the outside participants and requests their confirmation on the correctness of the information. In this design, the agent is programmed to communicate with the automatic online services of the third parties. The services can be provided by an online application or other agents communicated by standardized

communicating protocol, such as agent communication language (ACL), or Simple Agent Communication Protocol (SACP) (agent No. 4 in figure 1).

(4) Query data and take data to application agents for further confirmation: The mobile agent retrieves relevant information of the business operation in real-time and carries the information to other applications or devices for further confirmation. The example can be to examine the actual stock availability by communicating inventory management system (agent No. 5 in figure 1).

The functions of ASS include three modules: interface module, interpretation module and execution module, which are controlled by three software managers: Interface Manager, Procedures Manager, and Execution Manager (See Figure 1).

The Interface Manager manages the interfaces between the ASS and humane auditors. It is in charge of two functions: to capture required information for identification of specific audit procedure from audit requesters, and to present the execution result of the audit procedures to the audit requesters. (In adapting our study to a more automatic level, the entity for issuing audit procedures may not be limited to human auditors. An expert system which is capable to generate audit program is also technically feasible.) In order to pursue the first function, the Interface Manager needs a model to categorize the features of the audit procedures. As the conventional auditing professional has developed a framework for the organization of the comprehensive audit procedures, we will adopt the same framework. In the framework, the audit procedures are formulated as follows: `audit_cycle`, `audit_proc_type`, `audit_objective`, `sampling_method`, `timing`. That is, an audit procedure is specified by five procedure attributes: accounting cycle, audit procedure type, audit objectives, sampling method, and timing during the execution. The specific members (domain) of the first three properties are listed in table 1 (Alvin A. Arens, et al, 2003).

Regarding to the second function, two typical kinds of audit results will be presented: one is the error amount resulted from the execution of audit procedures, and another one is the alert to the auditors for the suspected operations that that are prone to be error or fraudulence. The Procedures Manager is responsible for mapping the audit procedures identified by Interface Manager into a detail plan for necessary and sufficient activities to collect the audit evidence. In this plan many information are required, including the audit evidences needed to collect, the mobile agents needed to dispatch, and the configuration of the mobile agents. Moreover, the information specified in the configuration of the mobile agent includes itinerary for the agents to travel, the information sources for mobile agents to access, and the computation procedure for the specified assignment

of the mobile agents. The Procedures Manager is supported by an activity database to generate suitable activities plan, where the activity database must be maintained by the auditing firm to adapt to the changes of audit requirement and client’s operation system.

Properties	Members (domain)
Accounting Cycle	Sales and collection cycle Payroll and personnel cycle Acquisition and payment cycle Inventory and warehousing cycle Capital acquisition and repayment cycle Cash balance cycle
Audit procedure type	Transaction-related audit objectives Balance-related audit objectives
Audit objective	Existence Completeness Accuracy Classification Properly classified Timing (Transaction-related audit objectives) Posting and summarization (Transaction-related audit objectives) Cutoff (Balance-related audit objectives) Detail Tie-In (Balance-related audit objectives) Realizable Value (Balance-related audit objectives) Rights and Obligations (Balance-related audit objectives) Presentation and disclosure (Balance-related audit objectives)

Table 1 Domain of property value for audit procedures

The Execution Manager is responsible for initiating and administrating mobile agents to perform required assignments. The mobile agents are initiated from the Aglet’s compiled java class code. Various kinds of abstract audit classes are pooled in the Agents pools and are accessed by Universal Resource Identifier from Internet. The pooling of the Aglets classes could be maintained as one of following three different ways: by local file directory (accessed by `file://c:/some/path/agent.class`), by Aglets server (accessed by `atp://some.server/path/agent.class`), or by http server (accessed by `http://www.some.host/path/agent.class`). The configuration of the agents is formatted as a file accessed by the agents as an input. When the mobile agents complete their assignment outside the ASS, they come back to ASS and transfer their results to Interface Manager. Then Interface Manager consolidates the audit result and formats the report to the audit requester.

3. CONCLUSION

In this study, we did not try to develop a comprehensive

system for implementing continuous auditing, which involve large efforts on the integration of several available computerized audit assist tools and a continuous audit model to seamlessly automate the whole audit process. In stead, our elaboration focuses on proposing a feasible and promising model for collection of audit evidence with continuous timeframe. Reviewing on literature for the proposal of continuous auditing, several models have been proposed. But each proposal has its limitations in practical implementation. In fact, auditors have problems to implement EAM models independently and have difficulty to embed the modules into clients' system after system being built. Similarly, ADM has its constraints on collecting limited types of audit evidence. Fortunately, the agent-based continuous auditing model can provide auditors to have an auditing system independently on developing enterprise system. Moreover, the installation of the system can be done even the system is running and evolving.

REFERENCE:

1. Alexander Kogan, Ephraim F. Sudit, Mikios A. Vasarhelyi. "Continuous Online Auditing: A Program of Research." *Journal of Information Systems*, 1999, 13(2), pp. 87-103.
2. Alfonso Fuggetta, Gian Pietro Picco, and Vigna Giovanni. "Understanding Code Mobility." *IEEE Transactions on Software Engineering*, 1998, 24(5), pp. 342-61.
3. Alvin A. Arens, Randal J. Elder, Mark S. Beasley. "Essentials of Auditing and Assurance Services: An Integrated Approach." Upper Saddle River, N.J: Prentice Hall, 2003.
4. Anuj K. Jain, Manuel Aparico IV, and Munindar P. Singh. "Agents for Process Coherence in Virtual Enterprise." *Communication of the ACM*, 1999, 42(3), pp. 62-69.
5. Danny B. Lange and Mitsuru Oshima. *Programming and Deploying Java Mobile Agents with Aglets*. Reading, Massachusetts: Addison-Wesley, 1998.
6. Harvey S. Koch. "Online Computer Auditing through Continuous and Intermittent Simulation." *MIS Quarterly*, 1981, 5(1), pp. 29-41.
7. Jon Woodroof and DeWayne Searcy. "Continuous Audit Implications of Internet Technology: Triggering Agents over the Web in the Domain of Debt Covenant Compliance," *the 34th Hawaii International Conference on System Sciences*. Tennessee University. TN USA: IEEE, 2001.
8. Miklos A. Vasarhelyi and Fern B. Halper. "The Continuous Audit of Online Systems." *Auditing: A Journal of Practice & Theory*, 1991, 10(1), pp. 110-25.
9. Robert K. Elliott. "Twenty-First Century Assurance." *Auditing: A Journal of Practice & Theory*, 2002, 21(1), pp. 139-46.
10. Roger Debreceeny, Glen L. Gray, Wai-Lum Tham, Kay-Yiong Goh, and Puay-Ling Tang. "The Development of Embedded Audit Modules to Support Continuous Monitoring in the Electronic Commerce Environment." *International Journal of Auditing*, 2003, 7, pp. 169-85.
11. S. Michael Groomer and Uday S. Murthy. "Continuous Auditing of Database Applications: An Embedded Audit Module Approach." *Journal of Information Systems*, 1989, 3(2), pp. 53-69.
12. Timon C. Du, Eldon Y. Li, and An-Pin Chang "Mobile Agents in Distributed Network Management." *Communications of the ACM*, 2003, 46(7), pp. 127-32.
13. Timon C. Du, Eldon Y. Li, and Eric Wei. "Mobile Agents for a Brokering Service in the Electronic Marketplace." *available online, Decision Support Systems*, 2004.
14. Tuomas Sandholm and Qianbo Huai. "Nomad: Mobile Agent System for an Internet-Based Auction House." *IEEE Internet Computing*, 2000, 4(2), pp. 80-86.
15. Yining Chen. "Continuous Auditing Using Strategic-Systems Approach." *Internal Auditing*, 2004, 19(3), pp. 31-36.
16. Zabihollah Rezaee, Ahmad Sharbatoghlie, Rick Elam, Peter L. McMickle. "Continuous Auditing: Building Automated Auditing Capability." *Auditing: A Journal of Practice & Theory*, 2002, 21(1), pp. 147-63.
17. Zabihollah Rezaee, Rick Elam, and Ahmad Sharbatoghlie. "Continuous Auditing: The Audit of the Future." *Managerial Auditing Journal*, 2001, 16(3), pp. 150-58.

Jurisdiction and Consumer Contract in E-Business

Zheng Tang

University of Birmingham, School of Law, Birmingham, B15 2TT, UK

ABSTRACT

The use of cyberspace increases tremendously in the past 20 years. An increasing number of businesses and consumers worldwide have accessed the Internet and there is few doubt on its growing significance for world trade. The unique features of E-business lead to the uncertainty and insecurity for the consumers as well as the companies, and incur great strikes to the traditional jurisdiction rules. It seems to be an exaggeration to argue that it would require totally new set of jurisdiction rules in E-business. It is, however, inevitable to reconsider appropriateness of the existing rules in some respects, including the reconstruction or reform of the resent rules.

This article will analyse the current jurisdiction doctrines to see how they work in the context of E-business and whether they advance the cause of consumers and the improvement of business. Specially, the article will study the Brussels Regulation as the recent most important and influential innovation trying to regulation jurisdiction on consumer contract in E-business. General suggestions will be given as to the future development to resolve this problem.

Keywords: Jurisdiction, Consumer Contract, E-Business, Brussels Regulation

1. INTRODUCTION

In the dawn of the new digital economy, “electronic business” shows its popularity in the wide world. Recent developments in the Internet communication technology have made it possible to sell goods and provide services through the Internet. The efficiency and convenience it provides encourage the consumers to take part in the new transaction model despite some concerns and doubts still remain. A non-bordered global market thus has been formed in the cyberspace, and an electronic business-to-consumer relationship has been built internationally, and at the same time, “inter-legally”. Jurisdiction problems thus arise as to which court will have the jurisdiction over the dispute on business-to-consumer contract in E-business.

To try to regulate activities carried out on the Internet by using the existing private international law rules is to put new wine into old bottle. The specific characteristics of the Internet pose uncertainty and lack of predictability to both the businesses and the consumers. When the consumer surfs on the Internet, he might easily unaware that he goes “abroad” and has transaction with a foreign company. The business also cannot identify the real origin of the consumer, thus submit itself to a potential court he has never expected. The present international trend on this problem is trying to protect the consumers by providing the court of the consumer’s domestic/resident country jurisdiction. This protective policy will subject the Internet businesses to the world-wide jurisdiction, which will greatly increase the cost and risk for doing E-business. The businesses may limit their markets much more than they would if the resolution of disputes were more predictable; correspondingly, the consumers may be frustrated because attractive products or services, or more competitive prices for certain goods or services are denied to them simply on the basis of their residence.

The net result of the situations is to the detriment of both the consumers and the businesses, and the whole promotion of consumer-orientated E-business.

This article will try to study the jurisdiction issues on consumer contract in E-business. It focuses on answering the following question: why the traditional jurisdiction rules do not work perfectly in E-business? What is the new development in this problem, and how is the effect of these new innovations, especially the Brussels Regulation? What are the possible suggestions for the future development and legislation on jurisdiction on electronic consumer contract?

2. THE DILEMMA OF TRADITIONAL JURISDICTION RULES IN CONSUMER ORIENTATED E-BUSINESS

Non-territoriality, non-discrimination access, intermediary involvement, and non-identity are some basic features of the Internet – the carrier of electronic business, which bring dilemma to the old doctrines of jurisdiction. The most basic and important feature of the Internet is its non-territoriality. The Internet is “a network of networks”, which is accessible from any computer locating anywhere in the world providing it is connected with the Internet Service Provider (ISP). The Internet communication thus breaks the territorial boundaries between the states in the physical world and makes distance disappear. As US District Court Judge Gertner stated that “The Internet has no territorial boundaries... as far as the Internet is concerned, not only is there perhaps ‘no there there’, the ‘there’ is everywhere there is Internet access”. [1] Secondly, the Internet is a decentralised system. There is no central or hierarchic controller in this system, which means the business that intends to establish a website dealing with the consumers only in a limited area cannot prevent the consumers worldwide accessing this information.

Thirdly, the Internet cannot run without the participation of intermediaries, which are the organisations whose services are used to facilitate a transaction between communicating parties. The intermediaries are usually not liable for the activities carried out by the Internet users. The location of the server thus often has rare or fortuitous relations with the liable online activities. Finally, the Internet technology enables users to communicate through the Internet anonymously, which blurs the information about the user's identity and location. However, in the traditional jurisdiction doctrines, such information may be important or even essential for determine the jurisdiction. These features challenge the function of the traditional jurisdiction rules in E-business.

2.1 Presence

The traditional practice in common law countries is to base the jurisdiction on the presence of the parties. If the defendant is present in the territory of a state, this state will have power to exercise jurisdiction over this person. The reason for presence basis is jurisdiction depends on physical control.[2] If the forum country has, directly or indirectly, an effective hold over the parties, it is practical for the forum to assume jurisdiction. In traditional English common law, an English court is competent to try action provided only the defendant has been served with a claim form. As to a company defendant, the English court can claim jurisdiction if this company registered in England, or outside England but has a business presence in England. [3] Other countries influenced by English law, such as Australia, Canada, New Zealand, Israel, and the United States, etc., all adopt this theory and base jurisdiction on the service of a claim form on the defendant.

Jurisdiction based on presence has faced challenges in E-business. The concept of "presence" has been questioned in the Internet age. As for individuals, a person might have some activities through the Internet and have actual connection with certain countries, which would make these countries appropriate fora. However, the person can just carry on these activities through the Internet and never be physically present in these countries. In order to settle this problem, it has been claimed that an individual is virtually "present" in the territory of one country if he appears on one website located within this country's territory. However, in this case, without the knowledge of the location of the server, the Internet user will hardly predict which jurisdiction his online action might bring himself into. This theory has a backward effect for the consumer, that it will make the business' home country in which the business locates its website server a competent forum.

A company can only have business presence in one country, either by its agent, its branch, or its place of business. However, in E-business, a company can get benefit from a foreign market only by establishing a

business website, without setting up any agent, branch, place of business or other physical establishment. Thus, no jurisdiction can be claimed based on presence principle. It has been suggested that the business website can be regarded as a business branch, or business establishment. If the website is accessible in one country, it can be regarded as a business presence in this country. The dilemma occurs that the business may thus be potentially subjected to all the jurisdictions in the world where the website can be accessed. Great uncertainty and unpredictability again exist. The presence principle thus turns more and more unpractical in the Internet age.

2.2 Residence/Habitual Residence And Domicile

Civil law tradition has based jurisdiction on residence/domicile of the parties. As the centre of the party's social activities, the domicile or resident country should be the most convenient forum for the concerned party. Subject one party to the judicial system of the party's habitual residence/domicile will satisfy the party's reasonable expectation in most cases. Further, a person will inevitably have close connect with his domicile/habitual residence, to subject a person to the close connected court will be appropriate and natural. However, it is not the case in E-business. Some activities carried out online may have very little connection with the businesses' residences/domiciles. For example, a business has established a website aiming at foreign markets only. In this case, for disputes arising over the business activities carried out through the website, it is unreasonable for the targeted court to decline its jurisdiction for the reason of the party has no residence/domicile in this country.

Another practical advantage of using residence/domicile to decide is certainty and predictability it provides. If the jurisdiction can be decided according to the parties involved, once the identities of the parties are settled, the probable jurisdictions are determined. However, in E-business, it is not the case. The Internet user is almost impossible to be identified. Without knowing anything about the true identity of the other party, the E-business participant cannot get any hint of where is the other party's domicile/habitual residence, and will not know in advance what is the probable jurisdiction this transaction may bring him into. Instead of certainty, residence and domicile bring unpredictability in E-business.

2.3 Nationality

Principle of nationality is another traditional doctrine to determine jurisdiction. The reason for this principle is based on personal bond between a party and a country. Use nationality as a basis for jurisdiction has been widely adopted by France and other Latin language countries influenced by France law.[4] However, this principle also cannot be efficient in E-business area.

Same as residence/domicile principle, when the Internet user is totally blind to the true identity of the other party, he will not know the other party's nationality, and will not have a sound prediction of the probable legal result of his activity. This principle has already lost its meaning in the virtual world.

2.4 Parties' Autonomy

Compared with the above doctrines, principle of parties' autonomy is a comparatively new principle in contractual area. Contractual parties can make agreement to determine which court may hear the dispute. This principle can help to protect parties' justified expectations and provide adequate certainty and predictability to foresee the legal result of a concerned transaction. These advantages make it a principle welcomed in the modern commercial world.

However, this principle has its limitations especially in E-business. The click-wrap contract is widely used in E-business. In this contract, the website provider, always the business, will display the terms and conditions of the agreement on the computer screen, and ask the other party, always the consumer, to click the button "I agree", in order to access this website or continue the purchase procedure. In a click-wrap contract, choice-of-court clauses are usually provided.[5] Although the click-wrap contract might be sound in efficiency, it is questionable whether it is also sound in fairness and justice. Without negotiating about the relevant rights and obligation, the consumer has to accept all the terms of the contract in order to make a purchase. The businesses thus have the advantage to choose the most convenient court to them, which might be vexatious or oppressive to the consumers. The result is unjust and unfair to one party with the benefit of another. This disadvantage might be detriment to the consumer's confidence to participate online transaction. Thus, the application of principle of parties' autonomy in E-business is also limited.

2.5 Place Of Performance

Place of performance has also been accepted as a basis to determine jurisdiction in commercial transactions.[6] However, it is quite difficult to determine where is the place of performance in E-business. Suppose an English company established a website based on the server in Japan, a Chinese consumer purchase an E-book through the website and directly download it to his computer. The consumer's ISP is located in Texas. The procedure of delivery is that: the product has been sent from the server, to the consumer's ISP, which will send it to the consumer's personal computer. Which country, Japan, China, or Texas, should be the place of performance? By holding the place of the performance is where the product is sent from, Japan will be the place of performance. However, Japan has little relation with both the parties and the transactions. Further, it is also

possible for the company to move the website into other server. The place of performance will change accordingly. If we take the place where the product has been received by the consumer as the place of performance, the remained question is whether the place of the consumer's ISP or the place of the consumer's computer can be the place of performance. Uncertainty still exists in this case.

2.6 Effect Has Been Caused In The State

Some country also bases their jurisdictions on the effect caused in their territories. Even if the party is outside the state's territory and has done actions elsewhere, if these actions cause effect in the state, this state will have jurisdiction.[7] Problem here is that by recognising this principle, a state greatly broadens its jurisdiction, and this broadness may cause uncertainty to the parties. Especially in E-business, a website accessible in one country will more or less cause some effect in this country. For example, an English company may just establishes a website to sell certain digital products to the consumers within Europe. The consumers in Texas who can access this website try to purchase some products online. It is reasonable to say the action of establishing website and uploading it has caused effect in Texas. However, is it fair enough to subject the company to the government of the Texas court? If so, an Internet company may be subjected to the jurisdiction all over the world. Criteria as to what kind of effect it must be have to be set up to take use of this principle.

3. BRUSSELS INNOVATION – DOSE IT WORK?

As mentioned above, the traditional jurisdiction rules have been questioned in E-business. It is an urgent problem to establish new approaches for the further development of consumer-orientated E-business. Many important international organisations have carried out certain work on this issue. EU Brussels Regulation might be the most influential and important innovation at present. Brussels Regulation re-organises and reforms some of the traditional jurisdiction rules and tries to make the variation efficient for the development of E-business.

Brussels Regulation has adopted domicile principle as the basic jurisdiction rule. It provides consumers the right of home forum action both as claimants and as defendants against the businesses. Parties' autonomy, as a general principle in private international law in contract, has been greatly limited in its efficiency and enforceability for consumer protection purpose. Brussels Regulation has also set up some conditions to keep balance between consumer protection and business promotion. It requires the businesses to pursue commercial or professional activities in the Member State of the consumer's domicile or, by any means, directs such activities to that Member in order to trigger the consumer protective jurisdiction provisions.[8] The

requirement in the Brussels Regulation is different from its previous precedent Brussels Convention, which requires the conclusion of the contract was preceded by a specific invitation addressed to the consumer or by advertising carried out in the consumer's domicile country and provided that the consumer took in that state the steps necessary for the conclusion of the contract itself.[9] The Brussels Regulation broadens the concept of "targeting" and removes the requirement of concluding contract in the consumer's home country. However, it is questionable whether the reform will achieve the goal of providing certainty and promoting E-business.

3.1 Targeting

First, is it required for the business to "target" or "aim specifically at" the jurisdiction in question in order to trigger the protective provisions? If we accept the requirement of targeting, problems will arise as to how to determine "targeting", especially in E-business. Brussels Convention has adopted the requirement of "targeting" by asking for "the conclusion of the contract was preceded by a specific invitation addressed to him (consumer) or by advertising" in the consumer's domicile.[10] This requirement has been criticized for it precludes where business can reasonably foresee the advertisement will reach consumer habitual resident/domicile in certain countries and will gladly accept the benefit thereof although the business may have had no specific intention to do business in such countries.[11] This difficulty is more obvious in E-business, for E-business brings worldwide consumers to the transactions without the necessary intentions of the businesses.

More problems will arise in E-business as to what the "special invitation" or "advertising" may be, and whether they can show the businesses' predict. The businesses generally will adopt the promotion method through email or website. The Email can be regarded as a special invitation sent to the consumers, and according to Brussels Convention, if the special invitation has been addressed to the consumer, it is enough to trigger the protective provision. The problem here is that it is not sure whether this "special invitation" by email is directed to the consumer's domicile/habitual residence.[12] For example, the business might send the consumer promotion email without knowing the actual residence of the consumer and it is hard to say the email can show his expectation to be governed by certain jurisdiction. Further, an Email is usually received and stored in the server, and then downloaded by the consumer. It is also possible to regard the place of the server as where the special invitation has been sent, and where the business is targeting. Even if we regard the place where the consumer accesses this email as the place being "targeted at", problem will arise when the consumer download and read the email in the place other than his residence.

More difficulties will arise about "advertising". Products promotion in the website can be regarded as advertising. According to Brussels Convention, the advertising in a website which is not really designed for the consumers in a particular jurisdiction is excluded from the scope of the consumer protection provision.[13] However, as the Internet is an open regime and accessible everywhere, the consumers outside the scope of the targeted countries have the equal opportunity to access the website and perform online purchase. The consumers online will not clearly know where this advertisement aims at, even if the business does give notice in the website that the products provided are for certain territories only, the consumers might be totally unaware of this announcement. In this case, is it fair to deprive these consumers protection from foreign litigation?

However, if we abolish the requirement of "targeting", it is also questionable. Since a website is accessible worldwide, and presently people have no truly efficient technologies to regulate its extension, a website with advertising "aiming at" a particular country will inevitably be viewed by the consumers from territories beyond the expected country. The potential result of refusing the requirement of "targeting" will subject the business to the actions in the unpredicted jurisdictions all over the world. For example, Brussels Regulation has made a somewhat radical reform in this point by providing "... the contract has been concluded with a person who *pursues* commercial or professional activities in the Member State of the consumer's domicile or, *by any means, directs* such activities to that Member State".[14] It is quite unclear what is the precise and authorized meaning of "pursue" and "direct". However, at least two points are certain. One is that the provision has been reformed purposely to include E-business, for "any means" will necessarily contain the means of Internet.[15] The other is the provision gets rid of the requirement that the concerned business activities must "aiming specifically at the country", which broadens the nature and scope of activity which can fall in the scope of the protective provision. Although the Commission in proposal for Brussels Regulation has stated that "a consumer simply had knowledge of a service or possibility of buying goods via a passive website accessible in his country of domicile will not trigger the protective jurisdiction",[16] it does not exclude the consumers who had more than simple knowledge of products information, to trigger the protective jurisdiction. For example, this provision enables the consumers to take action in their domiciles against foreign businesses, if they conclude contract via the active or interactive website which has been designed for special area outside the consumers domiciles. Suppose an English consumer accessed a French company's website, which was designed to sell music to the French consumers only. Although the French company made the statement to limit the service

area within France, the consumer did not notice it and purchased some music by providing credit card information and downloaded the music. Since present technology cannot efficiently prohibit this download, the company could do nothing to terminate the transaction but had to be potentially subject to the forum without previous expectation. That is the reason why many scholars and legal professionals hold the idea that this change goes too far to the development of online business.[17]

3.2 Necessary Steps To Conclude The Contract In The Consumer's Home Country

Secondly, is it required the contract must be concluded within the state of consumer's domicile/habitual residence? Brussels Convention has considered this issue and required that the consumer must have taken in the country of his domicile "the steps necessary for the conclusion of the contracts". However, in E-Commerce, it is clear the place of conclusion of contract becomes somewhat irrelevant to the matter. Without requirement of personal presence, a consumer can conclude contracts everywhere in the world. To the business, an English consumer concluding the contract online in his home country will have no difference if he concluding the contract in Japan. On the other hand, the business concluding contract online is not necessary to have a sound knowledge about the consumer's *concurrent* location, for it is quite often to be fortuitous without any connection with the consumer's domicile/habitual residence. In the above example, is it just to exclude the English consumer from the protective jurisdiction regime for the necessary steps to conclude the contract has been taken in Japan instead of England? If the invitation was sent individually to the consumer, or the advertising was designed aiming at the English market, is it reasonable to treat the consumer differently just because he happened to make this purchase outside England? Further, how about the business company has already known the domicile of the consumer before concluding contract? The fact that the consumer did not take necessary steps in his domicile will not damage businesses' reasonable expectation.

Furthermore, doubts might arise as to how to determine the places where the consumer has carried out the steps leading to the conclusion of the contract in E-Commerce. The character of the Internet makes it possible for "the steps necessary for the conclusion of the contract" to be taken in different States. For example, the consumer might begin the order in one country, continue negotiation in another, and complete it in the third country. Further, from the technical point of view, in a click-wrap contract, the contract is concluded in the server, instead of the place of consumer's location even if the consumer does click the bottom indicating his agreement in his domicile/habitual residence. In this case, the business can easily escape the protective provision by choosing the server located outside the

consumer's home countries. Even if we take the view that the contract is concluded in the consumer's location at the time of conclusion, it is for the consumer to prove that he has taken necessary steps to conclude the contract in his domicile in order to trigger the protective provision. The problem is how can the consumer prove the contract is concluded in his home other than in a foreign country?

By noticing these difficulties, Brussels Regulation has abandoned this requirement, but the reform has also been taken into question. It is claimed if the contract has been concluded totally outside the consumer's domicile/habitual residence, how can it be for the contract to have sufficient connection with the consumer's domicile/habitual residence? And how can it be just and fair for a court does not have sufficient connection with the contract to assert special jurisdiction over the case? Further, if the contract is concluded outside the consumer's home, the business will have no reason to expect the possible result of the transaction will subject himself to the unpredicted jurisdiction. For example, if the consumer performs the purchase in the business' domestic country, as to the particular contract, it is not clear from doubt that the consumer's unilateral behaviour is sufficient to bring the business to a foreign jurisdiction.[18] The language in the Brussels Regulation is so broad that it even can subject defendant company to the protective provision even if the contract were not concluded via the Internet but that nevertheless are related to and fall in the scope of the foreign defendant's website.[19] For example, a French company established active website with attractive advertisement in it to sell wine in France. This wine has never been put into the market outside France. An English consumer accessed this website and purchased the products in France during his one week travel. According to the provision, the company did direct "commercial or professional activities" to England "by any means", for the advertisement is accessible in England. However, is it reasonable and fair to subject the France Company to the English forum concerning lawsuit brought relevant to this wine?

4. CONCLUSION

Generally speaking, present jurisdiction rules, including some innovation tailoring the traditional jurisdiction rules to suit the development of E-business, are considered not satisfactory enough. The Brussels Regulation provides more certainty and security for the consumers, but may discourage the promotion of the business. Some undefined concepts, such as "direct to", "pursue" have been questioned and creates unpredictability to the parties. More work has to be done in this question. Although no systematic and detailed resolution recently, some general suggestions can be made to make this issue more clear:

First of all, certainty and predictability has to be guaranteed to both the businesses and the consumers. In this sense, if the traditional jurisdiction doctrines are adopted, further work has to be done to give precise definitions to some basic concepts in E-business.

Secondly, as unequal bargain power exists between the consumers and the businesses, more concerns have to be given for consumer protection. Restrictions should be set to regulation the effect of choice of forum clause in click-wrap contract.

Thirdly, in order to promote E-business, reasonable expectation of the businesses has to be protected. It is not sound to subject business to the jurisdiction of consumer's home country in any case. Some conditions have to be established to prevent over-protective solution for consumers.

Fourthly, as the technology is continuing growing, E-business in the future might rely on some technique other than the Internet. Thus, the resolution must be technology neutral so that it does not discriminate between different technologies.

REFERENCE

- [1] *Digital Equipment Corporation v Altavista Technology, Inc.* 960 F. Supp. 456 (D. Mass., 1997)
- [2] *McDonald v Mabee*, 243 U.S. 90, (1917), where Justice Holmes held: "The foundation of jurisdiction is physical power..."
- [3] CPR 6.5(6) (UK)
- [4] Civil Code, art 14, 15 (France)
- [5] See generally Matthew Burnstein, "A Global Network In a Compartmentalised Legal Environment", in Katharina Boele-Woelki and Catherine Kessedjian (eds.) *Internet: Which Court Decides? Which Law Applies?*, at p31-34
- [6] For example, see EC Regulation on Jurisdiction and the Recognition and Enforcement of Judgments in Civil and Commercial Matters, No 44/2001 of 22 Dec. 2000, (Brussels Regulation hereafter), art 5(1)
- [7] For example, see Restatement of Conflict of Laws, Chapter 3, § 37, § 50
- [8] Brussels Regulation, art 15.1(c)
- [9] Brussels Convention, art 13.3
- [10] Brussels Convention, art 13(3)(a)
- [11] See C. G. J. Morse, "Consumer Contracts, Employment Contracts and The Rome Convention", 41 ICLQ 3, January 1992, 1-21, at p7
- [12] See generally, Reinhard Schu, "Consumer Protection and Private International Law in Internet Contract", part C.I.1, Available from <http://ruessmann.jura.uni-sb.de/rw20/people/rschu/public/essay.htm>
- [13] Brussels Convention art 13.3(a)
- [14] Brussels Regulation art 15 (1)(c)
- [15] European Commission, Explanatory Memorandum to the Proposal for a Council Regulation on Jurisdiction, COM (1999) 348 of 14 July 1999.
- [16] EU Commission, Proposal for a Council Regulation (EC) on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters, Brussels, 14.07.1999, COM(1999) 348 final, 99/0154 (CNS)
- [17] See generally, Joakim ST Qren, "International Jurisdiction Over Consumer Contracts In E-Europe", ICLQ vol 52, July 2003, 665-696; Hague Conference, Draft Convention on Jurisdiction and Foreign Judgments in Civil and Commercial Matters, art 7(1)(a) has adopted a similar approach
- [18] For example, in several states of USA, in cases where the contract was unilaterally brought by the consumer to the consumer's forum, the court might deny the jurisdiction. See *World-Wide Volkswagen Corp. v Woodson*, 444 U.S. 286 (1980); *West Am. Ins. Co. v. Westin, Inc.*, 337 N.W.2d 676 (Minn. 1983); *Markby v. St. Anthony Hosp. Sys.*, 647 P. 2d 1068 (Wyo. 1982)
- [19] Mark Powell & Peter Turner-Kerr, "Putting the E-In Brussels and Rome", 1999 *International Company and Commercial Law Review*, 361-365, at 363; Marco Berliri & Markus Heyder, "Jurisdiction in Cyberspace: EU Jurisdiction Over the Rest of the World?", *Electronic Commerce & Law Report*, Vol. 6, No. 32, 885-889, August 15, 2001, at 887

Jurisdiction over E Commerce

Bill Loving, Xianping Liu

Idaho State University, Pocatello, Idaho

ABSTRACT

Expansion of business onto the Internet crosses national borders and creates uncertainty as to the propriety of jurisdiction by the courts of nations and their internal political units. Analysis of jurisdiction in United States law reveals that traditional rules provide guidance for resolving the question of constitutional jurisdiction in cases that cross geopolitical lines.

1. INTRODUCTION

Web commerce provides opportunities for international trading at low cost. Maintenance of Web sites and use of banner ads and search engines allow companies to enter highly desired United States markets and reach consumers who would otherwise be out of reach.

While both government and private businesses seek to take advantage of the annual estimated \$110 billion in U.S. imports,¹ this expanded commercial presence through the Internet means additional risk as American consumers and competitors turn to United States courts for redress of commercial injury.

The question of forum selection and jurisdiction has long been a central one in civil litigation in the United States.² Where can an aggrieved party seek redress: Where the injury took place, where the plaintiff resides or where the defendant lives? The question is complicated by the nature of e-commerce which bypasses national borders. For many Internet business models, the nationality of the consumer is relevant only in terms of exchange rates. That changes with the filing of a civil lawsuit. Then, the location of the forum in which the case is to be decided will be of great importance.

2. TRADITIONAL JURISDICTION ANALYSIS

¹<http://www.usashow.net/bmsc1.asp?id=381>

²Jurisdiction can be found as either general jurisdiction or specific jurisdiction. In general jurisdiction, the defendant will have an ongoing relationship and general activity in the state that is unrelated to the controversy at hand. See, *Helicopteros Nacionales de Colombia, S.A. v. Hall*, 104 S.Ct. 1868, (1984). An airline that operates in a state would be subject to general jurisdiction on the basis of its ongoing business presence. Specific jurisdiction arises when a defendant has sufficient contacts with the forum state that exercise of jurisdiction will not offend traditional notions of fair play and substantial justice. See, *International Shoe Co. v. State of Washington, Office of Unemployment*, 66 S.Ct. 154 (1945).

At the heart of jurisdiction is the due process clause of the Fourteenth Amendment of the United States Constitution. That clause has placed limits on which courts a defendant may be brought to answer for civil wrongs. That power of a court, tested in the enforcement of its judgments, is decided by the application of the Fourteenth Amendment, specifically the due process clause.³

The Supreme Court has dealt with the issue of defendant conduct and jurisdiction in three significant pre-Internet cases: *International Shoe Co. v. State of Washington, Office of Unemployment*, 66 S.Ct. 154 (1945); *World-Wide Volkswagen Corp. v. Woodson*, 100 S.Ct. 559 (1980); and *Burger King Corp. v. Rudzewicz*, 105 S.Ct. 2174 (1985).

International Shoe focused on the activities of an out-of-state corporation. *International Shoe* had between 11 and 13 salesmen who exhibited samples of the company shoes and solicited Washington residents to place orders with company headquarters in Missouri. The state of Washington sued to force the company to pay into the state unemployment fund.

The case reached the U.S. Supreme Court where *International Shoe* argued that due process restrictions would not allow the Washington courts exercise jurisdiction. Chief Justice Harlan Stone delivered the opinion affirming that *International Shoe* was subject to the power of that state courts. The key was the conduct of the shoe company in the state.⁴

³U.S.C.A. Amendment XIV II persons born or naturalized in the United States, and subject to the jurisdiction thereof, are citizens of the United States and of the State wherein they reside. No State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any State deprive any person of life, liberty or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.

⁴*International Shoe Co. v. State of Washington, Office of Unemployment*, 66 S.Ct. 154, at 159-160, (1945).

It is evident that the criteria by which we mark the boundary line between those activities which justify the subjection of a corporation to suit, and those which do not cannot be simply mechanical or quantitative.

Whether due process is satisfied must depend rather on the quality and nature of the activity in relation to the fair and orderly administration of the laws which it was the purpose of the due process clause to insure. That clause does not contemplate that a state may make binding a judgment in personam against an individual or corporate defendant with which the state has no contacts, ties or relations.

As a general proposition, Stone wrote, single or irregular activity by a corporation agents in a state did not make it liable to suit there. But International Shoe had done considerable business in Washington and, in so doing, benefited from the state⁵

[T]o the extent that a corporation exercises the privilege of conducting activities within a state, it enjoys the benefits and protection of the laws of that state. The exercise of that privilege may give rise to obligations; and, so far as those obligations arise out of or are connected with the activities within the state, a procedure which requires the corporation to respond to a suit brought to enforce them can, in most instances, hardly be said to be undue.

World-Wide Volkswagen Corp. v. Woodson, 100 S.Ct. 559 (1980) helped to refine the scope of the application of jurisdiction. In Volkswagen, a family purchased a car⁶ from a dealership that operated in New York. A year later, the family was involved in an accident in Oklahoma while en route to Arizona. The family brought a products-liability suit in Oklahoma. The Oklahoma Supreme Court ruled that the family could sue in an Oklahoma state court. It reasoned:⁷

The evidence presented below demonstrated that goods sold and distributed by the petitioners were used in the State of Oklahoma, and under the facts we believe it reasonable to infer, given the retail value of the automobile, that the petitioners derive substantial income from automobiles which from time to time are used in the State of Oklahoma.

Further, the Oklahoma Supreme Court reasoned, World-Wide could reasonably anticipate being brought into an

⁵Ibid. At 160.

⁶The car happened to be an Audi. The family was Harry and Kaye Robinson and their two children.

⁷ Quoted in World-Wide Volkswagen v. Woodson, 585 P.2d 351, 354 (Okla. 1978).

Oklahoma court.⁸ In the case before us, the product being sold and distributed by the petitioners is by its very design and purpose so mobile that petitioners can foresee its possible use in Oklahoma. World-Wide appealed to the U.S. Supreme Court and, in an opinion by Justice Byron White, the Court reversed. Justice White wrote that World-Wide had no contacts with Oklahoma, certainly not enough to give Oklahoma courts jurisdiction over the company.⁹ Here, there is a total absence in the record of those affiliating circumstances that are a necessary predicate to any exercise of state-court jurisdiction.

Justice White disposed of the Oklahoma Supreme Court analysis of the foreseeability of World-Wide being subjected to Oklahoma jurisdiction. Justice White acknowledged that World-Wide might foresee one of its cars being driven to Oklahoma and that such a car might be involved in an accident. But that was not the proper analysis of foreseeability, White wrote.¹⁰

The foreseeability that is critical to due process analysis is not the mere likelihood that a product will find its way into the forum state, but rather that the defendant conduct and connection with the forum are such that he should reasonably anticipate being haled into court there.

The mere fact that World-Wide injected a product into the stream of commerce that wound up in Oklahoma was not enough to create jurisdiction even though there was a claim that the product caused injury. In terms of libel, this case appears to foreclose most, if not all, jurisdictions outside the home state of the defendant publisher.

The Supreme Court fourth case in the quartet of jurisdiction decisions is Burger King v. Rudzewicz.¹¹ That case, involving a contract interpretation dispute between a fast-food franchisee and franchising company is relevant to the extent that it deals with the question of injuries that arise out of actions purposefully directed at forum residents.

Burger King sued Rudzewicz, the franchisee, in a Florida court. Rudzewicz opposed the exercise of jurisdiction and won in the 11th Circuit. Justice William Brennan delivered the opinion for the Court which reversed the 11th Circuit panel finding that jurisdiction was appropriate for the court in Florida.¹²

⁸Ibid.

⁹World-Wide Volkswagen Corp. v. Woodson, 100 S.Ct. 559, 566, (1980).

¹⁰Ibid.

¹¹Burger King v. Rudzewicz, 105 S.Ct. 2174 (1985).

¹²Ibid. At 2186.

Eschewing the option of operating an independent local enterprise Rudzewicz deliberately reach[ed] out beyond Michigan and negotiated with a Florida corporation for the purchase of a long-term franchise and the manifold benefits that would derive from affiliation with a nationwide organization.

When Rudzewicz failed to meet the terms of the contract, he caused foreseeable injuries to Burger King. Those injuries reasonably would lead to an accounting for the harm, Justice Brennan wrote. Rudzewicz availment of Florida laws before the relationship soured and his purposeful acts directed at Florida were enough to find jurisdiction.

3. APPLICATION OF FORUM ANALYSIS IN INTERNET COMMERCE

A case involving two United States companies from different states provides some direction to the question of jurisdiction over E Commerce. Zippo Mfg. v. Zippo Dot Com was a state law trademark dilution case brought by the Pennsylvania light maker against a California Internet company that maintained a Web site and provided an Internet news service.¹³ The Internet company did not have any offices or employees in Pennsylvania and so moved to dismiss the case because the Pennsylvania court lacked jurisdiction.

U.S. District Judge Sean McLaughlin began his analysis by citing the traditional three-part test for jurisdiction from International Shoe and Burger King: (1) sufficient minimum contacts in the forum state, (2) whether the claim arose from those contacts and (3) whether the exercise of jurisdiction would be reasonable.

He then turned to the changes that had taken place with respect to jurisdiction, running from physical to virtual presence.¹⁴ The Internet makes it possible to conduct business throughout the world entirely from a desktop. Judge McLaughlin then suggested a structure for deciding when jurisdiction clearly is and is not proper based on the quality and kinds of activity conducted over the Internet.¹⁵

This sliding scale is consistent with well developed personal jurisdiction principles. At one end of the spectrum are situations where a defendant clearly does business over the Internet. If the defendant enters into contracts with residents of a foreign jurisdiction that

involve the knowing and repeated transmission of computer files over the Internet, personal jurisdiction is proper. *E.g. CompuServe, Inc. v. Patterson*, 89 F.3d 1257 (6th Cir.1996). At the opposite end are situations where a defendant has simply posted information on an Internet Web site which is accessible to users in foreign jurisdictions. A passive Web site that does little more than make information available to those who are interested in it is not grounds for the exercise personal jurisdiction. *E.g. Bensusan Restaurant Corp., v. King*, 937 F.Supp. 295 (S.D.N.Y.1996).

Judge McLaughlin put interactive Web sites, sites that allow users to exchange information with the host, in a middle ground. The test for those sites would be¹⁶ the level of interactivity and commercial nature of the exchange of information that occurs on the Web site. *E.g. Maritz, Inc. v. Cybergold, Inc.*, 947 F.Supp. 1328 (E.D.Mo.1996).

Applying that test to Zippo Dot Com, Judge McLaughlin noted that Zippo had some 3,000 Pennsylvania subscribers to its news service. While instructive, Judge McLaughlin's analysis is little more than an application of the principles enunciated in International Shoe and Burger King. Even Burger King, which predates the Internet phenomenon, recognized the power of interstate communication to confer jurisdiction.¹⁷

Although territorial presence frequently will enhance a potential defendant's affiliation with a State and reinforce the reasonable foreseeability of suit there, it is an inescapable fact of modern commercial life that a substantial amount of business is transacted solely by mail and wire communications across state lines, thus obviating the need for physical presence within a State in which business is conducted. So long as a commercial actor's efforts are "purposefully directed" toward residents of another State, we have consistently rejected the notion that an absence of physical contacts can defeat personal jurisdiction there.

A case involving international parties followed a different approach. In *Hy Cite Corp. V. Badbusinessbureau.com, L.L.C.*,¹⁸ a U.S. district court in Wisconsin was called on to determine whether it could exercise jurisdiction in a dispute between an American company and a West Indian concern. In that case, the West Indian company operated a Web site called the Rip-Off Report, which featured critical statements about Hy Cite sued alleging unfair competition,

¹³Zippo Mfg. Co. v. Zippo Dom Com, Inc., 952 F.Supp. 1119 (W.D.Pa. 1997).

¹⁴Ibid. At 1123.

¹⁵Ibid. At 1124.

¹⁶Ibid.

¹⁷Burger King Corp. v. Rudzewicz, 105 S.Ct. 2174, 2184 (1985).

¹⁸Hy Cite Corp. V. Badbusinessbuearu.com. L.L.C., 297 F.Supp. 2d 1154 (W.D.Wisc. 2004).

false advertising, disparagement and trademark infringement claims under both state and federal law.¹⁹

Badbusinessbureau.com replied with a motion to dismiss for lack of personal jurisdiction. Judge Barbara Crabb, chief judge for the judicial district, concluded that the court did not have jurisdiction over badbusinessbureau.com either through general personal jurisdiction or specific personal jurisdiction. Judge Crabb cited the American constitutional requirements that a defendant have minimum contact in the forum state and, citing *International Shoe* and other federal cases, noted that those contacts must be purposeful and not random, isolated or fortuitous.²⁰

But where Judge McLaughlin in *Zippo* found jurisdiction based on a Web presence and the level of interactivity of the defendant business, Judge Crabb reached an opposite conclusion in *Hy Cite*. Judge Crabb rejected the *Zippo* approach on two bases: (1) Use of the interactivity test, and, (2) The lack of legal authority to craft a mechanical test for determining jurisdiction. Judge Crabb said it was not clear why interactivity should be a determining factor for jurisdiction.²¹

As even courts adopting the *Zippo* test have recognized, a court cannot determine whether personal jurisdiction is appropriate simply by deciding whether a website is "passive" or "interactive" (assuming that websites can be readily classified into one category or the other). And, Judge Crabb noted, operators of passive Web sites have been subjected to jurisdiction of courts when those sites have been used to harm plaintiffs intentionally.²²

4. CONCLUSION

The risk of lawsuits and the question of where those suits can be brought will profoundly affect e-commerce. A developing model sends advertising messages to Web users based on information they provide, either directly or through analysis of their site visitation. This advertising is not targeted to particular states, but rather looks to individual consumers and is based on a market for information. The Web site is not interacting with the residents of a particular state within the borders of the United States, but rather is seeking all persons interested in travel. The fact that a Web visitor is from a particular state in a particular country, becomes fortuitous and

happenstance contacts should not be enough to maintain jurisdiction.²³

Where, on the other hand, a business uses the Web to target consumers in a particular geographic or political entity, then jurisdiction should be proper in that foreign forum. The result from that system is foreseeability for businesses engaged in e-commerce that allows them to prepare to deal with litigants in foreign forums.

TABLE OF CASES AND AUTHORITIES

- [1] *Burger King Corp. v. Rudzewicz*, 105 S.Ct. 2174, 2184 (1985).
- [2] *Helicopteros Nacionales de Colombia, S.A. v. Hall*, 104 S.Ct. 1868, (1984).
- [3] *Hy Cite Corp. V. Badbusinessbuearu.com. L.L.C.*, 297 F.Supp. 2d 1154 (W.D.Wisc. 2004).
- [4] *International Shoe Co. v. State of Washington, Office of Unemployment*, 66 S.Ct. 154 (1945).
- [5] *Keeton v. Hustler Magazine, Inc.*, 465 U.S. 770, 774, 104 S.Ct. 1473 (1984).
- [6] *Noonan v. Winston Co.*, 135 F.3d 85, 92, (1st Cir. 1998).
- [7] *Panavision International, LP v. Toeppen*, 141 F.3d 1316, 1322 (9th Cir.1998).
- [8] U.S.C.A. Amendment XIV.
- [9] *World-Wide Volkswagen Corp. v. Woodson*, 100 S.Ct. 559, 566, (1980).
- [10] *Zippo Mfg. Co. v. Zippo Dom Com, Inc.*, 952 F.Supp. 1119 (W.D.Pa. 1997)

¹⁹ Ibid.

²⁰ Ibid. At 1158, citing *Keeton v. Hustler Magazine, Inc.*, 465 U.S. 770, 774, 104 S.Ct. 1473 (1984).

²¹ Ibid. At 1160.

²² Ibid. Citing *Panavision International, LP v. Toeppen*, 141 F.3d 1316, 1322 (9th Cir.1998).

²³ *Noonan v. Winston Co.*, 135 F.3d 85, 92, (1st Cir. 1998).

On the Novel Network Forensics Perspective of Enhanced E-Business Security

Wei Ren^{1,2}

¹ School of Information, Zhongnan University of Economics and Law,
Wuluo Road 114, Wuhan 430064, China

² School of Computer, Huazhong University of Science and Technology
renw@public.wh.hb.cn

ABSTRACT

E-business security is crucial to the development of e-business. Due to the complexity and characteristics of e-business security, the current approaches for security focus on preventing the network intrusion or misusing in advanced and seldom concern of the forensics data requiring for the investigation after the network attack or fraud. We discuss the method for resolving the problem of the e-business security from the different side of view - network forensics approaches - from the thinking of the active protection or defense for the e-business security, which can also improve the ability of emergence response and incident investigation for e-business security. It is also for the first time to systematically discuss the network forensics evidence source, network forensics principles, network forensics functions and network forensics techniques.

Keywords: e-business security, network security, digital forensics, network forensics

1. INTRODUCTION

Internet-based business operations offer many benefits, but also bring a broadened range of risks, some of them unprecedented. The actual and perceived lack of system security and reliability are significant deterrents to the rapid growth of the digital economy.

Daily progress is being made in reducing network risks through a variety of software patches, cryptographic algorithms and security tools. These efforts major focus on the prevention of the network intrusion, but always cannot eventually avoid the risk of the network misuse and fraud. We always pay more attention to the passive defensive measures of the e-business system, but ignore the computer evidence acquiring and forensics analysis after the attacks.

Although to some degree the legislation is relative later than the issues of computer and economics crime, once the system is compromised and leads to a large amount of economic lose of the enterprise, to obtain the computer evidence will become more and more important. On the other hand, forensics analysis of the system can also improve the protective and detective ability of network security system. To solve the puzzle of e-business security, we need different approaches to enhance the investigation of the network attack. Network forensics technology can be used for that purpose.

Judd Robbins, a prominent computer forensics investigator, defines computer forensics as “the application of computer investigation and analysis techniques in the interests of determining potential legal evidence.” [1] Other experts, believing computer forensics has evolved into a science, define computer forensic science as “the science of acquiring, preserving,

retrieving, and presenting data that has been processed electronically and stored on computer media.”

The term network forensics is commonly used to describe the task of analyzing information collected on active networks from various intrusion detection, auditing, and monitoring capabilities for the purpose of protection. The monitoring and analysis of data from live systems and networks will become essential to law enforcement as caseloads increase and juridical boundaries blur. [3,4,5,6,7]

In the First Digital Forensic Research Workshop, researchers give the definition of Network Forensics, [2] that is: The use of scientifically proven techniques to collect, fuse, identify, examine, correlate, analyze, and document digital evidence from multiple, actively processing and transmitting digital sources for the purpose of uncovering facts related to the planned intent, or measured success of unauthorized activities meant to disrupt, corrupt, and or compromise system components as well as providing information to assist in response to or recovery from these activities.

For the purpose of the network forensics, we always need the toolkits to capture the network traffic fully. There are many toolkits for building network traffic analysis and statistical event records [8,9,10,11,12]. They often use a promiscuous packet interface to pass visible traffic into an internally decision engine which discloses the content of the packets and counting them into statistical data and logging key details into backend disks.

After obtaining the network traffic data, forensics analysis is needed. Data mining techniques can be used for mining stream data or email contents [13,14,15]. Utilizing artificial intelligent approaches to identify

special features [16], IP trace back approaches [17,18] to the attack origin identification and mapping topology approaches for the possible location of the attack origin [19,20,21,22,23,24,25].

Current incident investigation mostly focus on the after attack data analysis. Seldom discuss the active investigative approaches in advance for potential risk and the techniques of the speedup of the emergence response time.

In this paper, we discuss the e-business security from a different point of view. Network forensics system that implement to capture the attacker's behavior and log them for the future analysis and investigation.

The remaining of the paper is organized as follows: First, Section 2 characterizes e-business security issues. Current solutions for implementing e-business security are discussed in Section 3. Section 4 details the discussion of network forensics approaches for e-business security. We give the conclusion and look ahead in Section 5.

2. E-BUSINESS SECURITY ISSUES

2.1 Types of Typical Attacks

There are many types of typical attacks that e-business corporation have to face and consider. The most common types are listed below:

Distributed Denial of Service (DDOS): This type of attack is often used when other protections have provided adequate security to the network. When such protections have denied attackers access, such attackers may resort to denying authorized users access to the network by overloading and hence crippling the network such that its performance significantly degrades or ceases to function altogether.

Viruses: This type of attack is often distributed via email attachment and often infects large numbers of customers and may be created itself replication. Viruses, once activated, may destroy information; provide future improper access to a network.

Data destruction: Improper access is gained and an entity's information is improperly changed or destroyed. Physical perimeter penetration: It is unauthorized accessing to a user's facility or network.

Password cracking: lists of the most used passwords are tried as a means of unauthorized access to another's network. Numerous cracker, hacker, web sites post lists of the most often used passwords.

Screen emulators: This is where low level access is gained to a network and a screen emulator is placed on the access server that brings up a false screen that emulates the proper login screen. This false screen asks for the users login and password and the brings up a screen that states "login incorrect, please try again." Actually the login was correct and the false screen

emulation program has captured another user's correct login and password.

Social engineering: This attack relies on the element of human weakness in protecting access information.

Other attacks that require more sophistication, such as cryptanalysis, man in the middle attacks, Trojan horses, IP hijacking, IP spoofing, sniffing, masquerading, reverse engineering and steganography or covert channels.

2.2 Objectives of e-Business Security

While the list of actual attacking manifestation is long, conceptually, they break down to a few categories. These are spoofing, unauthorized disclosure, unauthorized action, and data alteration. A well-planned information security strategy will address all these areas.

E-business security concerns also fall into four main categories: loss of data integrity; loss of data privacy; loss of service; and loss of control. Responding to these concerns requires an integrated and effective information security policy. In conducting e-business, every organization ought to be able to:

Positively identify or confirm the identity of the party they are dealing with on the other end of the transaction.

Determine that the activities being engaged in by an individual is commensurate with the level of authorization assigned to the individual.

Confirm the action taken by the individual and be able to prove to a third party that the entity did in fact perform the action.

To protect information from being altered either in storage or in transit.

Be certain that only authorized entities have access to information.

Ensure that every component of the e-business infrastructure is available when needed.

Be capable of generating an audit trail for verification of transactions.

Effective information security policy must have the following six objectives: privacy and confidentiality; integrity; availability; legitimate use (identification, authentication, and authorization); auditing or traceability; and non-repudiation. If these objectives could be achieved, it would alleviate most of the information security concerns or improve the e-business security circumstance.

3. CURRENT SOLUTIONS FOR E-BUSINESS SECURITY

3.1 Network Level Security

Network level security provides protection against attackers who attempt to deny service to legitimate users by gaining control of machines or resources within a

private network. The most common way to protect private networks that are connected to the Internet from these kinds of attacks is firewall technology. A firewall is located at a network gateway server that protects the resources of a private network from users from other networks. It is a combination of hardware and software used to implement a security policy governing the network traffic between two or more networks. The network firewall is the primary line of defense against external threats to an organization's computer systems, networks, and critical information.

In case an attacker successfully penetrates the firewall, IDS (Intrusion Detection System) can be useful to minimize the risk that any damage can be done to the servers or network. There are two kinds of IDS. One is Host IDS, which deploy on the host or server. The other is Network IDS, which set up in the network and detect the abnormal network traffic.

3.2 System Level Security

System level security is the ability to utilize operating system functions and applications in combination with hardware architecture to help protect against corruption of service and control user access to system resources, such as files, programs, databases and so on. Some technologies to improve system level security: host operating system harden, unsafe processes or services disabled, anti-viruses programs, database system security, data backup and recovery, application security and so on.

3.3 Transaction Level Security

The actual act of completing transactions on the Internet depends on transaction level security. Transaction level security refers to the ability of two entities on the Internet to conduct a transaction privately and with authentication. In assessing e-business security, all the components of a transaction have to be considered: the client, transport channel, servers, operating system, applications and database components. Both the authentication of the component of the transaction is crucial, so it always needs the support of PKI system. This level of security provides a basis to enable the payment for goods and services to occur in privacy. Currently, the leading technologies for implementing transaction security are Secure Socket Layer (SSL) and Secure Electronic Transaction (SET).

3.4 Weakness of Current Solution of e-Business Security

Assuring the availability and security of the network is complex challenge. Historically enterprises employ a patchwork of nonintegrated security products that provide incomplete coverage. There are two weakness of existing solution to the e-business security.

One is individual point of products. Over the past decade, security analysts and network managers have come to rely on a multitude of specialized solutions that address specific points of the network. These include firewall, IDSs, server log files and vulnerability scanners. They are not built into a cooperative or integrated system. Once system is compromised, all products are unfortunately do not present a holistic picture of network activity because their initial purpose is protection but not forensics.

The other weakness is limited sampling. Current products can only capture limited quantity of network traffic. This creates a frustrating guessing puzzle for network managers or incident investigator who must interpret these ambiguous snapshots of the network traffic or attack behavior.

4. NETWORK FORENSICS FOR E-BUSINESS SECURITY

4.1 Network Forensics Evidence Source

Computer evidence must keep the criterion of the traditional evidence, such as confidential, authentication, integration, legitimated. Moreover, it has its own characteristics listed as follows:

Virtual: Computer evidence consists of the computer binary data and files in the computer or devices. They cannot be touched and looked without the processing of computer software.

Multiform: The appearances of the evidence are diversity because they can be stored with different format, such as the words, picture, wave sound and video. Some of them even can be interactive with the computer user. They can be stored in the hard disk, USB disk, compact disk and so on.

Fragile: The computer evidence is easy to destroy by the software or the human interference. It can be deleted, modified and counterfeited by the mistakes of investigators or the premeditation of criminals.

Instant: Some evidence is altered with the time, such as the network traffic packets, process in the host, open ports. They will not replay if investigators do not record them on the disk. The forensics system can record the dynamic data and save them to a safe disk, where other softwares cannot modify the data, so that it will keep the originality and integration of the evidence.

There are four types of data source for the computer evidence.

The data from host or server: It includes the operation system audit trail, system event log, application event log, alert log file, file MAC (Modify/Access/Create) timestamp, recovery data, system time, file slack, erased files, swap files, memory contents, program files. It is the traditional computer forensics data.

The data from network: It is the network traffic data packets. Because the network becomes more and more frequent channel to launch attacks, the data of the traffic

becomes also more and more important to forensics analysis. Especially for the network forensics system, the traffic is a key data source. However, the traffic data always has large volume, therefore the efficiency of dumping data is the major issue of the network forensics system.

The data from other security products: It includes firewall log, IDS log, access control system, router, network interface, PDA and so on. The data on these systems give some digest information of the intrusion and misuse.

The data from Internet: For the Internet forensics, the data from Internet can give some basic information for investigation, such as the personal data from the personal website or search engine, the content of chat room.

4.2 Network Forensics Principles

The evidence obtained after the network forensics process must keep the legal validity, so some principles must be satisfied in the network forensics procedure.

To keep the originality of the data: The analysis is not directly on the processing of the original data, but on the base of image mirror copy of the original data. The copy is a way of bit-to-bit copy and is the exactly copy of the original data. All the forensics analysis process on the copy version of data will keep the integration of original data and make the replay of analysis procedure possible.

To keep the result believable: It means the forensics software and hardware is safe and believable. The software to capture data, analysis data, process data, and display final evidence must be the authority software or satisfy the standard of the forensics procedure. Before the forensics standard becomes uniform and certificated software are available, the open source software may be a temporary way with believable.

To keep the procedure believable: Very step of processing action must keep the integration of the analysis data. The data signature or data digest approaches can be utilized for this purpose. Read-only property of files is not always believable because even read-only compact disks may also have data changes on special conditions.

To keep the integration in transportation: If the data need to be transported, the integration and confidentiality of the data must satisfy. Data signature also can be used in this occasion.

Documentation and monitoring: The whole procedure of the forensics must documentation, including the time, location, operator name, approaches, techniques, operations steps, processing results and so on. The procedure must be in the monitor of the third party or authority organization, and they must give the signature on the document of the result.

4.3 Network Forensics Functions

(1) Network investigating

Before the enough evidence is available, some investigation can be provided by the network forensics system. Search engineering tools is the fundamental program in the network forensics system suite or integrated into the system. Browser tools, ftp tools, email tools and other Internet tools are also needed.

(2) Network surveying

Some network survey tools are also included. The first is footprinting tools, such as whois, nslookup, traceroute. The second is scanning tools, such as nmap, Hping2, which can be added into the network forensics package or customized development. The third is enumeration tools used for netbios enumeration, snmp enumeration and active directory enumeration.

(3) Network traffic recording

Network traffic is fully dumped by the network forensics system, which can also filter the traffic according to the rules. Rules can be customized for different purpose.

(4) Data aggregation

Logging data from different location give different feedback of the attacking behavior. The analysis of the aggregation of the data sets, which are from multiple sources, such as firewalls, IDSes and sniffers, can build the chain of the clues and display the full scene of the crime. Network forensics system can aggregation the data and transform the data into a uniform data file or database.

(5) Future attacking pattern predicting

The hacker group always has some features, such as the types of attacking tools, the frequently utilizing techniques, their often steps and trace routes for intrusion. Therefore the network forensics system can use data mining approaches to discover the potential pattern of attacks and provide the function of predication.

(6) Anomaly pattern discovering

The log data in the forensics system can be mined for the anomaly pattern, which will also give a feedback to influence the setting of firewall rules and intrusion detective signatures.

4.4 Network Forensics Techniques

(1) Mapping topology

Building the topology database and IP location Mapping topology of the network may help to find fraud proxy server, ARP spoofing, or quicken the location of the attack origin.

(2) Honeypot/honeynet learning and collecting

Using the honeypot system and network forensics analysis, we can build a database to profile the blackhat, person or organization, such as the name, nickname, email address, home address, nationality, age and so on. We can store the IP address, blackhat techniques, tactics, motives and psychology in the database. We can use dig

tools to profile the main IP node domain name, topology of network or the location of the hackers. The data in the database can be update automatically and also keep the current data and old data for the future timeline analysis.

(3) TCP session replaying

To analysis the attack behavior by replay the attacking procedure. In the captured network traffic, unrelated packets appear in the order they were transmitted over the wire. Network forensics tools can reorganize the packets into individual transport-layer connections between machines. To reassemble the connections, more forensic details emerge.

(4) Protocol parsing

Protocol parsing and analysis is the major work of network forensics analysis. In the analysis, the POP3, HTTP, FTP and telnet protocols need to be paid more attention.

(5) Covert channel discovering

After the protocol parsing, we need to find the covert channel or data hiding in the traffic. Some attacker use steganography in the communication, it add the burden of the investigation.

(6) Potential pattern recognizing

Some artificial intelligence approaches can be used to forensics analysis. Two types of machine learning approaches can be built into network forensic systems: Artificial Neural Networks and Support Vector Machines. Since the ability to identify the important inputs and redundant inputs of a classifier leads directly to reduced size, faster training and possibly more accurate results, it is critical to be able to identify the important.

(7) Forensics data stream mining

We can also use some data mining approaches to network forensics analysis. Data mining generally refers to the process of extracting models from large stores of data. We choose several types of algorithms in our research: Classification, maps a data item into one of several predefined categories; Link analysis, determines relations between fields in the database. Finding out the correlations in forensics data will provide insight for discovering attack behavior quickly; Sequence analysis, models sequential patterns. These algorithms can help us understand the sequence of forensics events. These frequent event patterns are important elements of the behavior profile of a user or program.

(8) IP trace back to the attack origin

In the investigation we can use some methods to trace a steady stream of anonymous Internet packets back towards their source. These methods do not rely on knowledge or cooperation from intervening ISPs along the path. Sometimes tracing an attacking stream requires only a few minutes once the system is set up for a victim.

(9) Remote OS fingerprinting

Remote OS fingerprinting is always a technique on footprinting. It can obtain the general OS type of the target host. This is useful to estimate the experience level and the possible attack tools of the investigate

object. The result also can as a digital evidence for the future forensics.

(10) Remote network forensics

Remote network forensics is a program to capture the network traffic on the remote host. Always it is employed on the local area network forcedly or some key traffic center for capturing fully traffic that is used for the future forensics analysis.

5. CONCLUSION

Network forensics approaches for the e-business security can trace the behavior of the fraud in the e-business, discover the potential risk through the analysis the detail forensics data, quicken the speed of emergence response, enhance the ability of the rapid incident investigation, providing the evidence for the future legal action. The future work is the improvement of the dump performance of the network traffic in the network, development of more mining tools for the analysis of the forensics data.

ACKNOWLEDGEMENT

Thanks go to Dr. Jin Hai for his insightful discussion of the project and paper. We also thank the security team of the Cluster and Grid Computing Lab in School of Computer, Huazhong University of Science and Technology.

REFERENCES

- [1] Osles, L. "Computer forensics: The key to solving the crime", 2001.
- [2] Gary, P. (2001) "A Road Map for Digital Forensic Research", Technical Report DTRT0010-01, DFRWS, November 2001.
- [3] Corey, V.; Peterman, C.; Shearin, S.; Greenberg, M.S.; Van Bokkelen, J.; "Network forensics analysis", Internet Computing, IEEE, Volume: 6 Issue: 6, Nov.-Dec. 2002 Page(s): 60-66
- [4] Brian Carrier. "Defining Digital Forensics Examination and Analysis Tools". In Digital Research Workshop II, 2002
- [5] Mark Reith, Clint Carr, Gregg Gunsch, "An Examination of Digital Forensic Models", International Journal of Digital Evidence Fall 2002, Volume 1, Issue 3
- [6] J.Tan. "Forensic Readiness". In The CanSecWest Computer Security Conference, April 2001.
- [7] Yanet Manzano and Alec Yasinsac, "Policies to Enhance Computer and Network Forensics", The 2nd Annual IEEE Systems, Man, and Cybernetics Information Assurance Workshop, at the United States Military Academy, June 2001
- [8] Giovanni Vigna Andrew Mitchell, "Mnemosyne: Designing and Implementing Network Short-Term Memory",
- [9] S. Ioannidis, K. G. Anagnostakis, J. Ioannidis, and A. D. Keromytis. "xPF: packet filtering for lowcost

- network monitoring". In Proceedings of the IEEE Workshop on High-Performance Switching and Routing (HPSR), pages 121--126, May 2002.
- [10] S. McCanne and V. Jacobson. The BSD packet filter: A new architecture for user-level packet capture. In Proc. of the USENIX Technical Conf., Winter 1993
- [11] K. G. Anagnostakis, S. Ioannidis, S. Miltchev, and J. M. Smith. Practical network applications on a lightweight active management environment, In Proceedings of the 3rd International Working Conference on Active Networks (IWAN), pages 101-- 115, October 2001.
- [12] Fulvio Rizzo, Loris Degioanni, An Architecture for High Performance Network Analysis, Proceedings of the 6th IEEE Symposium on Computers and Communications (ISCC 2001), Hammamet, Tunisia, July 2001.
- [13] O. de Vel. "Mining e-mail authorship" In Proc. Workshop on Text Mining, ACM Discovery and Data Mining (KDD'2000).
- [14] W. Lee and S. J. Stolfo. Data mining approaches for intrusion detection. In Proceedings of the 7th USENIX Security Symposium, 1998.
- [15] W. Lee, S. J. Stolfo, and K. W. Mok. Mining in a data-flow environment: Experience in network intrusion detection. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-99), August 1999.
- [16] Srinivas Mukkamala & Andrew H. Sung, "Identifying Significant Features for Network Forensic Analysis Using Artificial Intelligent Techniques", International Journal of Digital Evidence., Volume 1, Issue 4, Winter 2003.
- [17] Stefan Savage, David Wetherall, Anna Karlin, and Tom Anderson. "Practical network support for ip traceback", In Proceedings of the 2000 ACM SIGCOMM Conference, August 2000, An early version of the paper appeared as techreport UW-CSE-00-02-01
- [18] Hal Burch and Bill Cheswick. "Tracing anonymous packets to their approximate source". In Proceedings of the USENIX Large Installation Systems Administration Conference, pages 319--327, New Orleans, USA, Decemeber 2000. USENIX.
- [19] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. "Network topology generators: Degree-based vs structural", In ACM SIGCOMM, August 2002.
- [20] Bill Cheswick, Hal Burch, Steve Branigan, "Mapping and Visualizing the Internet" ,USENIX Annual Conference, General Session - June 2000,
- [21] D. Magoni and J.J. Pansiot. "Analysis of the autonomous system network topology", ACM SIGCOMM Computer Communication Review, pages 26--37, July 2001.
- [22] Ramesh Govindan and Hongsuda Tangmunarunkit. "Heuristics for Internet Map Discovery", In Proceedings of the 2000 IEEE INFOCOM Conference, Tel Aviv, Israel, March 2000.
- [23] A. Lakhina, J. Byers, M. Crovella, and I. Matta, "On the Geographic Location of Internet Resources", Technical Report BUCS-TR-2002-015, Boston University, 2002.
- [24] C. Jin, Q. Chen, and S. Jamin. Inet: Internet Topology Generator. Technical Report Research Report CSE-TR-433-00, University of Michigan at Ann Arbor, 2000.
- [25] A. Medina, A. Lakhina, I. Matta, and J. Byers. BRITE: An Approach to Universal Topology Generation. In Proceedings of IEEE MASCOTS'01 August 2001..

Regulating Copyrights in Cyberspace

Sylvia Mercado Kierkegaard

kierkegaard@privat.dk

ABSTRACT

The explosive growth of the Internet has created critical new challenges to national and international intellectual property policies. Intellectual Property protection has become the most important issue in Cyberspace. When technological advances has outpaced the law, several problems arise which require an approach suitable to the digital age. Several international conventions have been enacted with principal objective of bringing world law into the digital world. The TRIPS Agreement and the WIPO Copyright Treaty were adapted, which require Member States to give protection to digital content providers against copyright infringements. However, as with the Berne Convention, TRIPS and WIPO do little to reduce substantial disharmonies in the substantive content of national copyright laws. The digital age has seen three controversial views on intellectual properties emerge. There are people who believe that intellectual property should be unprotected and unrestricted. On the opposing view, Intellectual Property Right (IPR) owners feel that the national governments need to pass and enforce laws to protect intellectual property. They claim that violation of intellectual property is inhibiting them from investing and making information more available in cyberspace. Others contend that traditional copyright law is able to deal effectively with digital copyright issues in cyberspace. At the same time, there is a need to maintain a balance between the rights of authors and the larger public interest. The core issue is how to protect copyrighted materials while at the same time serving the public's right for privacy, information access and dissemination.

Keywords: copyright, digital management, file-sharing, circumvention

1. INTRODUCTION

The advent of the Internet over the past 10 years has paralleled the emergence of globalization as a concept. It has become the principal means of global information. In a relatively short amount of time, the Internet has become a ubiquitous tool throughout the world. Net population is expected to hit 950 million by year 2005. Consumers are taking advantage of higher speed networks and new technologies to share and distribute information. The ease and speed in which contents are copied and distributed over the Internet have become a contentious issue with major companies, in particular the movie and recording industry, demanding stronger copyright regulation and enforcement. Concern over theft of intellectual property and the ensuing technological and financial losses have raised new clamours among legislators and Intellectual Property Rights (IPR) owners for a harmonized international intellectual property law to govern the Internet. IPR owners claim that violation of intellectual property is inhibiting them from investing and making information more available in cyberspace. Opponents of cyberspace regulation insist that there is little to fear on this issue as IPR owners could easily manage the problems through tighter technological safeguards and self-regulation. Traditional copyright law is able to deal effectively with digital copyright issues in cyberspace. Others contend that the current copyright law must be amended to adapt to the new technologies and to serve the public's right to information. At the heart of the opposing views is the issue of protecting ownership of copyrighted material, while at the same time facilitating greater ease of access and online content. This paper will examine the various issues and problems

surrounding copyright protection in Cyberspace and determine whether the current international conventions and enforcements are sufficiently applicable to the Internet regime.

2. COPYRIGHT

Copyrights are rights given to persons over what one has created. They usually give the creator an exclusive right over the use of his/her creation for a certain period of time. Copyrighted work can be a literary work, musical work, dramatic work, pantomime, choreographic work, pictorial work, graphic work, sculptural work, motion picture, audiovisual work, sound recording, architectural work, mask works fixed in semiconductor chip products, or a computer program. Also protected through copyright and related rights are the rights of performers, producers of phonograms (sound recordings) and broadcasting organizations. The main social purpose of protection of copyright and related rights is to encourage and reward creative work. The owner of a copyright has the right to exclude any other person from reproducing, preparing derivative works, distributing, performing, displaying, or using the work covered by copyright for a specific period of time. The rights of a copyright owner are subject to a number of qualifications such as fair use, compulsory licenses and substantiality.

3. INTERNATIONAL LEGAL FRAMEWORKS

Most countries are members of the Berne Convention for the Protection of Literary and Artistic Works (Berne Convention) which gives protection to works in countries of which one is not a citizen or national. It has

been signed by 156 countries since 1886. It is a multi-national agreement on copyrights which is administered by the WIPO, a specialized agency of the United Nations. In addition to proscribing discriminatory treatment of qualified foreign authors, Berne sets forth minimum standards of protection that must be accorded to those authors even if a country accords its own authors lesser protection. The provisions serve to protect the following author's rights: the right of reproduction including translations and adaptations, the right of dissemination to a non-material form including public performance, broadcasting (etc.), and the right to protect their moral interests. Under the Berne Convention, the following works may be protected: (i) both unpublished and published works of an author who is a national or resident of a country that is a member of these treaties; or (ii) published works, with permission, of an author who is not a national or resident of a country that is a member of these treaties (Art.5). In this case a work may be considered simultaneously published in several countries if it has been published in two or more Berne countries within 30 days of its first publication (Art.3). Copyright under the Berne Convention is automatic: no registration is required, nor is the inclusion of a copyright notice. The Berne Convention provided for a minimum term of copyright protection of the life of the author plus fifty years, but parties are free to provide longer terms of copyright protection. Indeed, the principal contribution of Berne to the international norms of copyright has been in persuading countries to agree on the principle of national treatment under which the works of foreign nationals receive the same level of protection as the country accords to its own nationals. Even though the Berne Convention has achieved some harmonization in national copyright laws by virtue of its establishment of minimum standards, it nonetheless defers to national traditions to a considerable degree. While national laws may be different in other countries, one is guaranteed at least certain protections. Unfortunately, the Berne Convention does not have the judicial nor police power to enforce the Convention. Neither does it have an effective means of calling signatory countries to account for non-compliance with Berne standards. It can, however, withhold copyright protection to authors from non-Berne countries and exert pressures on them to enact or amend their laws to conform to the same requirements as the Berne Convention.

The Universal Copyright Convention (UCC) provided an alternative mechanism by which states could enjoy international protection. It was adapted in 1952 under the auspices of UNESCO to cater to the US objections against the Berne Convention. The United States refused initially to become a party to the Berne Convention, since it would have required major changes in its copyright law (particularly with regard to moral rights and the registration of copyright works). Although it is not as exhaustive as Berne Convention, it requires formalities concerning deposits, registration and fee

payment. While the Berne Convention does not establish any formal requirements to benefit from the copyright protection other than having the author's name on the work, the UCC requires a copyright notice. It embodies the principle of national treatment, but not automatic protection. This notice should consist of the copyright symbol "©" accompanied by the year of first publication and the name of the copyright owner, all "placed in such a manner and location so as to give reasonable notice". Generally works are protected for a minimum of 25 years beyond the life of the author.

The World Intellectual Property Rights Organization (WIPO) has sought a leading role in the international copyright administration and harmonization. Since 1967, it has administered the Berne Convention. WIPO and the signatory nations to the Berne Convention have developed a new international treaty, known as the WIPO Copyright Treaty (1996), which extends the rights of authors in the digital era. 47 countries have signed the treaty. Explicit references to new technologies are found in Articles 4 and 5 which extend the copyright protection to computer programs and compilations of data. Article 7 creates rental rights in respect of computer programs. Articles 11 and 12 contain important provisions obligating member states to prevent circumvention of technological measures used to protect copyrighted works, and to prevent tampering with the integrity of copyright management information.

The Trade-Related Intellectual Property Rights (TRIPS) agreement appended to the General Agreement on Tariffs and Trade (GATT) went into effect on January 1, 1995 as part of the agreement that established the WTO. It was designed to remedy the perceived weaknesses in the Berne Convention and other multilateral intellectual property treaties. TRIP establishes minimum levels of protection that each government has to give to the intellectual property of fellow World Trade Organisation (WTO) members. The Agreement adds rights beyond those that it incorporates from Berne by requiring countries to make available effective remedies for copyright enforcement. By bringing copyright within the ambit of the WTO, it provides a mechanism for international enforcement through the imposition of trade sanctions against non-complying countries. The agreement covers five broad issues: how basic principles of the trading system and other international intellectual property agreements should be applied; how to give adequate protection to intellectual property rights; how countries should enforce those rights adequately in their own territories; and how to settle disputes on intellectual property between members of the WTO. Each of the main elements of protection is defined, namely the subject-matter to be protected, the rights to be conferred and permissible exceptions to those rights, and the minimum duration of protection. The Agreement sets these standards by requiring that the substantive obligations of the main conventions of the

WIPO, the Paris Convention for the Protection of Industrial Property and the Berne Convention must be complied with. All the main substantive provisions of these conventions, with the exception of the provisions of the Berne Convention on moral rights, are incorporated by reference and thus become obligations between TRIPS Member countries. Under Articles 10 and 11 of TRIPS, GATT countries are now bound to protect computer programs as literary works. It outlines how databases should be protected and expands international copyright rules to cover rental rights of authors of computer programs and producers of sound recordings. A similar exclusive right applies to films where commercial rental has led to widespread copying which is materially impairing the right of reproduction. It also requires performers to be given protection from unauthorized recording and broadcast of live performances (bootlegging). Producers of sound recordings must have the right to prevent the unauthorized reproduction of recordings for a period of 50 years. When there are trade disputes over intellectual property rights, the WTO's dispute settlement system is now available.

One hundred and forty seven (147) members of the WTO have accepted the terms of the Berne Convention as part of the trade-related aspects of intellectual property rights (TRIPS). Moreover, for the twenty five member states of the European Union, the rights are enshrined within Directive 29/2001/EC. However, as with the Berne Convention, TRIPS actually does very little to reduce substantial disharmonies in the substantive content of national copyright laws.

4. COPYRIGHT ISSUES

Many legal issues arise in cyberspace, but no other single area of law presents such a variety of interesting and diverse legal problems. This is true in no small part due to the fact that people all over the world are increasingly becoming connected via the global telecommunications networks. As Internet usage grows, new legal questions associated with the technology continue to surface. Currently, all content available in cyberspace is protected under copyright law. The reality is that digital communications and the digitization of information of all types make the infringement of intellectual property rights, particularly copyrights, easier than ever before. Existing intellectual property doctrines may not provide the level of control that content-providers would like. Copyright law may not be able to protect authors' rights in cyberspace, in the same manner as it has in the physical world. The problem may not be the law, but the difficulty in enforcing copyright protection in cyberspace and the public perception that information on the Internet is public domain. Cyberspace is not as constrained by national boundaries as the physical world. The old legal system, developed during a time when borders were important,

is probably inadequate to deal with the borderless nature of the Internet and the rapid technological changes.

4.1 Enforcement

According to the OECD's 2004 Information Technology Outlook, file-sharing networks are being used to freely download more movies, games and software than music. The number of people logged on simultaneously to popular file sharing networks approached close to 10 million in April 2004. The Motion Picture Association claims that it is losing hundreds of millions of dollars to Internet piracy. The origin of movies circulating on the Internet varies. Some are advance copies of blockbusters, apparently stolen from studios or otherwise leaked to the public before they are shown in theatres. The copies are converted into digital files and put on the Internet, or resold in the form of illegal DVDs and videocassettes. In other instances, people armed with digital video cameras make their way into an early screening of a movie, record the movie and turn it into a digital file. Once a movie has been released on the Internet, it spreads via peer-to-peer file-sharing services.

New studies show that contrary to protests from record labels, piracy is not responsible for the 15 percent drop in music sales in the past two years. According to Forrester Research, Inc., a research firm providing data and analysis that defines the impact of technology change on business, there is no evidence of decreased CD buying among frequent digital consumers. Times are tough on the music business, not because of downloading, but due to other factors such as economic recession, competition from videogames and DVD sales and the levies that are arbitrarily applied to music-playing equipment like MP3 players at wildly varying levels in different countries. It could also be that the introduction of CD anti-piracy methods has alienated the consumers. There is evidence that some of the protected CDs do not play correctly even on normal CD players and consumers complain that even when a CD is purchased, it cannot be played in the format of choice or the medium of choice.

Although the European Union (EU) has strict harmonised laws against piracy, the enforcement varies among Member States. In the first quarter of 2004, the Motion Picture Association sent out more than 350,000 cease and desist letters of which illegal downloads accounted for 90%. Seventy-seven % (271,000) of these letters were sent to Internet Service Providers (ISP) in the European region. The MPA uses tools to monitor and take evidence to sweep the Web for evidence that movies are illegally online and then asks Internet service providers to block access to illegal content. Many of the ISP providers in Europe have remained uncooperative and insist that anti-piracy groups present a valid search warrant before they can be allowed to access the user's file in order not to intrude on the

privacy of their customers and violate their trust.

The degree of enforcement of rules also varies from country to country, with Northern European countries more vigilant than their Eastern and Southern European counterparts. In the EU, the problem is not the lack of regulation, but the absence of uniform enforcement. Movies illegally downloaded from the Internet are hawked openly in the streets of downtown Madrid and restaurants, while such a brazen display of pirate copies will not be possible in Germany.

US authorities prosecute both downloading and uploading of files, but in Europe, the crackdown affects only uploaders. In April of 2003, three American college students became the first computer users who were forced to pay fines ranging from \$ 12,000 to \$15,000 by the Recording Industry Association of America (RIAA) for swapping unauthorized music online. They were sued by the RIAA for creating search engines on campus networks that made it easier to locate and share files that reside on others' computers, including term papers, research papers, photographs and MP3 music files. They settled their suits by agreeing to pay thousands of dollars over time. The targets were discovered by using exotic computer hardware called Spybots to scan publicly available peer-to-peer (P2P) networks and identify the ISP (Internet Service Provider) of each user. Under the Digital Millennium Copyright Act (DMCA), the RIAA can subpoena ISPs for each user's name, address and other personal information. This procedure requires only a court clerk's signature and need not come before a judge to be effective and the information can be used in collecting and identifying information of subscribers on the basis of mere suspicion. However, a recent decision by the U.S. Court of Appeals for the District of Columbia overturned a trial judge's decision to enforce copyright subpoenas, one of the most effective tools used by the recording industry. The court said the 1998 copyright law doesn't cover popular file-sharing networks used by tens of millions of Americans to download songs. Thus, the recording industry can't force Internet providers to identify music downloaders,

RIAA also went to court to shut down Napster, a peer-to-peer (P2P) file sharing service. Immediately, less centralized systems like KaZaA quickly replaced Napster. The industry has not so far persuaded the courts that these digital copying and sharing technologies are themselves "contributory" infringers of copyright. A US court has refused to order the shutdown of peer-to-peer file sharing services operated by Grokster and Streamcast Networks. Judge Stephen Wilson of U.S. District Court in Los Angeles in *MGM Studios, Inc. v Grokster Ltd* (Case Nr. CV 01- 8541) has effectively ruled that those who have no direct control of the use of their services - over the files swapped - cannot be held responsible for any misuse of those services. Without evidence of their active and

substantial contribution to the infringement, the file-trading services cannot be held liable.

If enforcement is left to the national states, IPR owners fear that piracy will continue unabated because of the laxity of authorities and corruption that has not been eradicated in some places. Infringement is possible from any corner of the globe. Countries that have not signed to the Berne convention are often referred to as copyright jungles, as there are often rampant copyright violations in those countries. The absence of any international convention on jurisdiction regarding the Internet makes it difficult to enforce copyright authority, even if the infringer has already been identified. The *Convention on Cybercrime* is the first international treaty which addresses this problem by requiring member countries to adopt similar criminal laws on hacking, copyright infringement, and computer related fraud. It also contains provisions on investigative powers and procedures, such as the search of computer networks and interception of communications. It also requires cross-border law enforcement cooperation in searches and seizures and extradition.

IPR owners also oppose self-regulatory schemes. There is no real protection because self-regulatory schemes cannot produce enforceable decisions. It can only work if it can rely on a government apparatus for enforcement. Even if the legislative rights of authors are expanded, the availability of technology that enables rapid and cheap copying of content in the digital world, as well as the global and borderless character of the Internet will render enforcement difficult outside the national boundaries of the country. Although copyright law attempts to dissuade future copyright violations, rampant piracy continues to plague the Internet and piracy of computer software has risen to an alarming level even in countries with strict intellectual property rights laws. It is apparent that the present laws are not producing their intended deterrent effects.

4.2 New Technology

Time barrier is becoming less of an issue as more people gain access to high-speed Internet connections. The software and video industry thought that slow download speeds would be a safe haven from non-commercial file-sharing, but that time is gone. Now, consumers with broadband can download a feature-length film in about six hours. The movie industry fears that advances in technology will significantly reduce that time. Current technology (most P2P file-sharing software) allows constant connections with other computers on the network as long as the software is running. Although file-sharing software is frequently associated commonly with illegal downloading of music and movies, it is a powerful technology that allows efficient distribution of legitimate files and data of all types. Record companies and movie studios may not willingly cede control to

users who have legitimately purchased their products. It will be difficult for them to lock down digital video so tightly. Someone will eventually find a way around it. The Internet abounds with freely available software that lets consumers circumvent copy-protection systems used on commercial DVDs.

4.3 Free Speech and Human Rights Principle

Copyright can also interfere with free flow of ideas, information and commerce. The Internet serves as a worldwide public commune, exposing the people of the world to a dizzying variety of ideas, expressions, cultures and creations in a manner so free and easy that it sparks an immense amount of additional creativity and innovation. The Internet's shelves will be empty unless individuals and institutions possessing relevant information place it in the Internet. To gurus of copyright free cyberspace, information should be free—that is, the public should have the right to share, enjoy, criticize, and build on the works of others in order to promote progress. One extreme view states that there are no property rights in cyberspace, and therefore, there is no need to obtain authorization to reproduce a copyrighted work. It is argued that by posting materials on the Internet, copyright owners are granting implied licenses to users.

Large corporate interests seek to narrow the freedom of the Internet for their own economic advantage. Many countries have adopted the Digital Rights Management, which control the accessing and use of creative materials in ways that are often inconsistent with a free and democratic copyright system. These systems are developed on the assumptions that IPR owners can exercise complete control over works in which they hold copyright. Therefore by attempting to wall off portions of cyberspace they are destroying the Internet's potential to foster democracy and economic growth worldwide. In print publication, no license is required to purchase a book, but in the digital world, organizations are imposing copyright license. This requirement prevents access to information and infringes on the individual's right to privacy as identification requirements are demanded on the users.

Mistaking the free distribution of content with the placement of intellectual property in the public domain is common among web users. If there is no copyright notice posted on the website, it is assumed that anyone can feel free to use it. But this argument can be fallacious if the same principle governing copyright of books is applied on the Internet. An exemption would be "fair use" for scientific and educational purposes. Use of copyrighted works or portions thereof, for any other purpose is not deemed fair use. Property holders can distribute their works freely while retaining their right of control over that work.

4.4 Unfair Protection of a few interest groups

Increased copyright protection and enforcement in their countries, as mandated by TRIPS and the Berne Convention, has mainly benefited the industrialized countries, in particular the United States. Multi-national enterprises from industrialized countries are the main producers of copyright-protected works, and the developing countries (third world) are primarily copyright users. The Berne Convention has been sharply criticized for hindering the development of developing countries by making it too restrictive for the latter to access the materials. Many claim that the Berne Convention is an artefact of the 18th century intended for another type of works and not for the Internet technology. Critics contend that Berne protects the interest of the industrialised countries by ensuring that the creators will be able to recoup their cost and prevent copyist from the developing countries from offering identical products at very low prices. This will enable the creators to charge prices for access to those works substantially greater than they could in a competitive market. They contend that this is economically inefficient, wasteful of social resources and deter progress. In order to make the Convention relevant to the needs of the developing countries, pundits are advocating the creation of exemptions to copyright restrictions, improved affordable access and assistance to the developing countries.

In the United States, the "Sonny Bono Copyright Term Extension Act" extended the term of copyright protection to nearly a century for corporations and even longer for many individuals and their heirs. A number of products – books, films and music have been given unreasonably longer copyright protection and prevents them from being made available in the Internet. The 1998 law that extended copyright protection to 95 years for most existing copyrights and 70 years after the death of the author for most new ones. The Act has been criticized for protecting private groups and condoning their rent-seeking behavior by granting them a monopoly in order to solidify their dominant position in a marketplace and exempting them from the free market pressures. A constitutional challenge to the Sonny Bono law was rejected by the Supreme Court in 2003 (*Eldred v. Ashcroft No. 01-618, No. 01-618*). The Court's decision implies that Congress can freeze the public domain indefinitely. Defenders of public interest have organized to lobby for changes in the copyright system which would allow classical and valuable works to be made available for the public interest.

The law has been formulated and drafted by legal experts who are heavily influenced by IPR owners. These industry groups are well organized and have strong lobby powers with legislators and bureaucrats, and very often the latter relies on their views for decision-making. The current legal and political regime is inclined to extending more rigid protection to IPR

owners and not the user's rights. Copyright protection of the Internet cannot favour one group over another. Unfortunately, the consuming public has not been consulted, mainly because they are disorganized and consist of individual users. Any framework convention needs to be based on agreements involving different constituencies. There is a need to revise the law to accommodate individual user's rights and to draw a line between legitimate copying and not – what should be allowed on the web and what is illegal. A decade ago, lawsuits and clarifications of copyright law resulted in "fair use doctrine, which permitted the limited reproduction of copyrighted material by journalists and scholars.

4.5 Difficulties in Legal Interpretation

Courts take a variety of approach in interpreting copyright law. In the light of furious litigation in the United States, the latter is seen as taking a stricter interpretation of the copyright law than its European counterparts. The Swedish MP3 case exemplifies the problems of applying the old intellectual property legislation to the digital media. The plaintiffs were a number of international record companies who claimed that the defendant, Mr. Olsson, made illegal copies of the CD's and then via hyperlinks from his own homepage made these copies available to the public. The court found that the hyperlinks on the defendant's homepage could not be regarded as a public display or performance. Neither was the hyperlink considered as "distribution of a work" based on the fact that Mr. Olsson did not produce the copies made available via his homepage because it was the Internet-users who downloaded the music file from the hyperlink. Although the putting up of a hyperlink is considered a public performance, the defendant was found not guilty because the Swedish Copyright Act exempts public performance of sound recordings from copyright.

The courts have not always been equal to the task of resolving copyright conflicts protection. In the US, there have already been many lawsuits involving the Digital Millennium Copyright Act (DMCA). In one early case, the federal government criminally prosecuted a company that created a device to decrypt electronic books. A jury eventually acquitted the company. But in another case, online journalists who distributed "DeCSS," a program for decrypting DVDs, were found to have violated the DMCA even though the program could be used in ways that would not infringe copyright.

There is a need for clarification and enforcement of copyright law. Court decisions have failed to settle various issues associated with copyright (such as hyper linking, circumvention, framing, and liability of access providers and users) and the question of how much sharing should be allowed or whether all of it should be stringently prosecuted as a violation of copyright law.

5. NEED FOR BALANCE

The Digital Millennium Copyright Act (DMCA), which was signed into law in 1998, created significant new rights for copyright owners. Whereas copyright law previously centered on the exclusive rights of copyright owners to perform certain acts with their works, this new legislation for the first time created a right to exclude access to copyrighted works. The DMCA prohibited the distribution of technology that circumvents the industry's electronic locks on books, films, articles, software, or songs, even though circumvention itself could contribute to research and technological innovation. The DMCA protects owners of copyrights, without regard to whether those owners were the creators of the protected work. It regulates technology that controls access to and use of works, rather than regulating the use itself. Its so-called "anti-circumvention" and "rights management" provisions facilitate copyright owners in preventing others from viewing their works and from performing acts that would previously have been allowed under the "fair use" exception to the copyright laws.

Copyright protection technology faces a legal challenge in France where a consumer association is filing damages in a legal action which contends that the copyright protection used by EMI label to prevent CDs from being pirated, stops consumers from making personal copies of their CDs - a privilege granted to French consumers by a 1985 law- and makes it impossible to play them on many car stereos, home stereos and personal computers

In the United States, a maker of software that enabled users to copy DVD's and computer games folded under the mounting weight of lawsuits filed by movie studios and video games producers. Unfavourable court rulings by three federal courts in 2004 assured the demise of 321 Studios. A federal judge in New York imposed a worldwide ban on the production and distribution of 321's Games X Copy software, which let users make what 321's Web site had called "a perfect backup copy of virtually any PC game." Hollywood and the computer-gaming companies accused 321 of violating the 1998 DMCA. The company's software was meant to let consumers make backup copies of their DVDs and computer games. Consumer advocates warn that the court decision restricts the rights of the consumers to make copies of their own legally obtained digital materials.

In contrast, a Norwegian Appeals Court (Case Nr. 02-507 M/94) upheld an earlier verdict that Jon Johansen, a 20-year old Norwegian man, had not broken the law by creating a system that could get around copy protection on DVDs. The ruling is a setback to anti-piracy efforts by the Hollywood studios. Mr Johansen, known as "*DVD Jon*" by the net community, created his program to watch films on a Linux-based

computer. He then posted the program onto the net in 1999. His software, called DeCSS, could decrypt disks by stripping the Content Scrambling System from DVDs. The US movie industry had accused DVD Jon of theft. But an Oslo court said in January 2003 that he was free to do what he wanted with DVDs he bought legally.

Song-swapping is not the only copyright infringement that the music industry is fighting. EMI music issued a cease-and-desist letter in February 2004 to a small-time record producer demanding that he halt distribution of a clever musical mix he had made. Producer Brian Burton's "*The Grey Album*" electronically combined sounds from the Beatles' recording commonly known as "*The White Album*" with rapper Jay-Z's "*The Black Album*", without seeking permission from the artists or their labels. More than 100 Web sites rallied to his defense by offering "*The Grey Album*" for unauthorized, free downloading as part of an ad-hoc protest.

6. CONCLUSION

Although the copyright law's penalty system attempts to dissuade future copyright violations, rampant piracy continues to plague the industries. It is apparent that the present statutory remedies are not producing their intended deterrent effects. Self-regulation and innovative new technology could be the answer.

The adaptation of IPR regimes to the digital economy is needed to protect rights embedded in the new technology developments. However, the new legal rules must take cognizance of the need to balance public interest, privacy, economics and copyright protection. Any international convention would have to define how the Internet is to be governed with consideration to the technological elements and the machineries for monitoring and enforcement compliance.

Since cyberspace is constantly and rapidly in transition, copyright law may not be able to catch-up with the speed of new technology. This suggests that copyright law may not be able to protect the author's rights in the digital world in the same manner as it has in the physical world. Because many countries prefer to preserve their own copyright tradition, it would be difficult to harmonize international law. IPR owners should find new technological remedies to assist the law. Any insufficiency in the law or in enforcement of authors' rights in cyberspace may be more than made up for by emerging technological means of protecting digital works. Adoption of technological means to protect works against unauthorized use or to track down infringements may, in fact, mean that authors' rights will become better protected in cyberspace than they have ever been in the physical world. Just as people use technology to infringe copyrights, people can use technology to stop it rather than relying on the threat of litigation. The marketplace already offers a variety of technology-based solutions to meet a variety of

copyright protections. Among the relevant emergent technologies are digital envelopes, encrypted signal streams, software metering schemes, digital watermarks, and copyright management information attached to digital copies of works.

There are legal file sharing websites which gives access to a vast catalogue for a monthly membership fee. Apple Computer's approach is to use the carrot instead of the stick. While recording labels' approach is to sue, Apple has set up the successful *iTunes* music store and has attracted people to that model of paying music legally. In the year since the introduction of *iTunes* in the United States, Apple has tweaked its system for managing music rights slightly - for example, by reducing the number of times a list of songs can be burned onto a CD, and increasing the number of computers that can play a copy of a purchased song.

The industry is being far too heavy handed by fining people hundreds of US dollars. The damages they seek are disproportionate to the cost of purchase of one song. One of the reasons people download music files is because of the high prices of CDs. Rather than fighting a losing battle, the IPR owners should be looking at ways of updating its failing business model. Record companies must be prepared to embrace the new technology, find out how people are using it, and work out how to profit from it. Instead of lobbying to preclude the use of MP3 files, they must ask themselves why customers are choosing to download files instead of legitimately purchasing the CDs. It is possible that people no longer want CDs or tapes, not only because of the cost implication. Compression breakthroughs have made it easy to quickly download and distribute music files. This distribution can allow consumers to discover and follow new bands and to meet other fans with shared interests. An MP3 can hold up to 10 archived albums and can be played in most modern car or stand alone stereos. It allows users to compile their own music play list, unlike purchasing legitimate CDs which often contain only few albums with significant appeal. It seems that record companies are having serious trouble when it comes to picking winning albums.

The economics of the Internet dictate that business must find a way to generate revenues without charging users for intellectual property. Any new regulation must be suitable and flexible to deal with the challenges posed by Internet technology and must balance the interest of the different stakeholders.

REFERENCES

- [1] OECD Information Technology Outlook 2004.
- [2] www.forrester.com
- [3] <http://www.technewsworld.com/story/35033.html>
- [4] Free Expression Policy Project, a think tank on artistic and intellectual freedom, www.fepproject.org

Research on Policies to Support Telecommuting in China

Peipei Wang, Pengzhu Zhang, Fangfang Xu

Aetna School of Management, Shanghai Jiaotong University, Shanghai 200052, China
{ wangpeipei, pzzhang, lotue }@sjtu.edu.cn

ABSTRACT

Telecommuting (or Telework) is now relaxing more and more workers from traditional 9-to-5 routines in many countries. In this paper, we do some research on China's actuality, and try to give out some reasonable suggestions to support Telecommuting, on the basis of learning from the developed countries.

Keywords: Telecommuting /Telework, advantages, government, policy

1. INTRODUCTION

Telecommuting (or Telework) is a work practice that entails working at least some of the time, often conceptualized as remote working supported by information and communication technologies (e.g. Gray et al. 1993).

Early definitions characterized telework as working away from a traditional office (Grant, 1985; Kelly, 1985) or at home (Regeneye, 1985). Later definitions were broadened to include the use of information and communications technologies (ICTs) (Holti and Stern, 1986; Oslon, 1988).

The emergency of the telework has various reasons. For example, Knowledge workers are separated from line-workers and needn't always stay in their offices. At the same time the rapid development of communication technologies also made it possible for the companies to implement telework, such as e-mail, faxes and messenger.

Globalization and the establishment of multinational organizations have increased the distance between physical flows (concerning production, distribution and consumption of products) and information management and are enlarging the distance between business branches, also between producers and buyers. Employees, who are not on the same worksite, perhaps in the different countries, cannot contact each other face to face everyday. All these call for a kind of more flexible and convenient work style--- Telework.

Also, telecommuting was driven by concerns about traffic congestion and pollution in densely polluted areas such as southern California and telecommuting was treated as a means to reduce the expense of maintaining office space. More recently, organizations have begun to view telecommuting as a tool to attract and retain top personnel in fields with short labor supplies.

In the 1980s, the first telecenters were established in Europe. Ten years later the emergence of the digital

economy vastly expanded the potential for performing jobs in telework arrangements. In the 1990s the U.S. federal government also instituted a government-wide telework pilot project, and many states and local governments, as well as private-sector corporations, instituted telework programs and published results from their experiences. Today, with the proliferation of technology in the workplace, telework is one of the fastest-growing policies promoted by savvy employers. The 2000 Telework America Research Survey estimates that the number of teleworkers in the U.S. has increased by 2.8 million since 1999, bringing the current total to 23.6 million. Gartner Group predicts that more than 137 million workers worldwide will be involved in some form of remote work by 2003.

In the Office of Personnel Management's (OPM) 2003 report to Congress on the status of telework in the federal government, the Director of OPM described telework as an "invaluable management tool which not only allows employees greater flexibility to balance their personal and professional duties, but also allows both management and employees to cope with the uncertainties of potential disruptions in the workplace, including terrorist threats."

In this paper, our main interest is to give some suggestion for government and local authorities to support telecommuting based on china actuality.

2. DIFFERENT TYPES OF TELECOMMUTING AND THEIR ADVANTAGES AND DISADVANTAGES

2.1 Different types of Telecommuting

A meaningful classification of telecommuting for the purposes of the present article is based on the geographical location of the workplace.

- Satellite work centers;
- Neighborhood centers;
- Work at home; and
- Mobile work.

Satellite work centers are self-contained organizational

units belonging to companies; it's in some sense a branch. The relocation of such units is made possible by the separation of functions in modern organizations and by the possibility of integrating shared facilities, such as databases, through information technologies. The advantage of such relocation exists in the possibility to reduce office costs in the headquarters, usually located in central areas of the cities, and in the reduction of commuting for employees, and also for the customers. It's equipped with office furniture and equipment provided by the firm, and administrative help may be available.

Neighborhood centers, different from satellite offices, are remote offices for employees from several organizations, which share space and equipment. The location of such offices is not driven by the need to realize economies of scale. Therefore employees can choose to work in the neighborhood centers that are closest to their homes. The employees are sufficiently independent in their work and the employers can rely on remote control and communicate with them. A network of neighborhood centers is also a good support for the communication between itinerant employees and their headquarters. Office suits may be furnished by the site owner or by each renting firms.

The establishment of neighborhood centers may be critical for the development of some suburban areas, especially those that are suffering problems of unemployment. In these areas a neighborhood center may support the diffusion of computer knowledge and promote new activities and latent potential.

Work at home refers to employees who work at home on a regular basis, though not necessarily every day. It is the most common interpretation of telecommuting. Teleworkers need communication medias such as telephone, electronic mail, personal computer links to office servers through ISDN lines, and fax machines.

Home Telework does not provide the social interaction that a satellite or neighborhood working center would theoretically provide, but could provide employees with extreme flexibility in schedule and lifestyle. It would allow employees to decide autonomously when and where they want to work and how they wish to accommodate childcare and other family duties. Work at home also offers new job opportunities to the elderly and the handicapped.

In contrast to the three above, mobile workers are frequently on the road, using communication technologies to work at home, in a car, in a plane, or at a hotel—communicating with the office as necessary from each location. It gives employees the chance to accommodate their personal/family life as well as work responsibilities. Therefore, mobile workers should have a good access to information networks anywhere.

2.2 Advantage and disadvantage of telecommuting

These alternative work forms bring both advantages and disadvantages to organizations, individuals, and society. We will discuss them on the assumption that the teleworkers is away from the office a significant percentage of working time. The impact may diminish if telework is just a small fraction of normal working time. First let's talk about the advantages.

From the organization's perspective, telecommuting provides an opportunity to improve workplace productivity. Also, it enables them to cut cost related to office space, as well as to address space constraint. Besides, it widens the talent pool available, improve recruitment and retention of employees by providing workers with flexibility that allows them to keep their jobs in the face of external demands or desires. Finally, the organizational advantages cover a wide spectrum, including behavioral outcomes, productivity, and legal issues.

From the employees' viewpoint, they will have greater job satisfaction and higher work performance ratings. First, it greatly reduces the time and money spent on commuting, eliminates the necessity to relocate. Second, it provides employees with a relatively distraction-free environment, greater schedule flexibility and autonomy. Third, they and their families can benefit more from the new work/family balance.

To the entire society (table1), telecommuting can decrease traffic congestion on strained highways and reduce automobile-related air pollution-although its impact may be negligible if only a few people telework. Further, teleworkers discourage community crime and spend more time building community ties.

On the other hand, several shortages may hamper the willingness to integrate telecommuting into traditional office environment.

To the organization, there are problems from two points. For managers, the main problem is their inability to physically observe their subordinates' performance. Secondly, the difficulty to create team synergy, because telecommuting will negatively affect the social network by disrupting teamwork and so on. The other issue concern developing and transmitting organizational culture, a virtual culture is needed and clear telecommuting policy guidelines should be established.

To the employees, probably the most commonly expressed challenge of telecommuting is overcoming the isolation caused by the separation of the telecommuter from the social network in the traditional workspace. Besides, teleworkers should overcome the challenges of maintaining access to resources at work, and they need to be more technically savvy than their

office peers, since support services are not accessible on-site.

To the society (table 1), telecommuting does pose some challenges. Telework may cause a gradual degradation of a person’s ability to interact with others in a personal,

intimate and human manner. This may at last change the nature of social intercourse. The main issue is to establish telework culture, zoning regulations that forbid business parcel delivery within residential neighborhoods.

Table1: Societal Advantages and Challenges of Telecommuting

	Advantages	Challenges
Home-Based Telecommuting	<ul style="list-style-type: none"> ●Less traffic congestion ●Less pollution ●Less neighborhood crime ●Greater community involvement 	<ul style="list-style-type: none"> ●Telework culture ●Loss of ability to interact with others
Satellite Office	<ul style="list-style-type: none"> ●Less traffic congestion ●Less pollution ●Greater community involvement 	
Neighborhood Work Center	<ul style="list-style-type: none"> ●Less traffic congestion ●Less pollution ●Greater community involvement 	
Mobile Work		<ul style="list-style-type: none"> ●Telework culture

Source: Nancy B. Kurland, Diane E. Bailey[15]

Table 2: Policies used in some developed countries and their Effects

Countries/ areas	Relative policies	Number of teleworkers	Effects
America Washington	The federal government and some of the largest businesses there are urging more of their employees to stay home.	About 15% of the region’s 2.6 million employees work from home, up from 250,000 five years ago. They telework at least 1 day/month in 2003, with most averaging 1.5 days/ week according to the Metropolitan Washington Council of Governments.	If the effort’s goal is met, 17,000 daily vehicle trips and 550,000 vehicle miles could be eliminated by June 30, 2005.
America	The American Disabilities Act (ADA); The Clean Air Act of 1990; The Family Leave Act of 1993.	Approximately 23.5 million teleworkers in the United States compared to 11.6 million in 1997.	The average business can save \$5,000 by reducing its workspace. (Said by Bob Smith, executive director of the International Telework Association and Council)
Toronto, Canada	A Toronto telework plan offering corporations free telework consulting and training help.		
Austin, Texas	Use a Web-based service to allow employees to telework at home, in order to improve poor air quality on July 2002.	Approximately 400 employees telework about one day a week through the program. The city has set a goal of 600 employees work from home one day/week by the end of the summer2003. Within two years, Austin officials want to see the total rise to 1,000.	Austin officials hope that telecommuting will help reduce the total vehicle miles traveled by its employees by 15 percent over the next five years
Malaysia	Amend the Employment Act 1955 to include new flexible employment regulations.	Prime Minister Datuk Seri Dr Mahathir Mohamad said the group’s involvement in the workforce in 1995 was 43.5 per cent and 45.7 per cent in 2003.	Enable women’s participation in the labor force and enable them to juggle between work and family responsibility.

3. TELECOMMUTING AND POLICIES SUPPORTING TELECOMMUTING IN SOME DEVELOPED COUNTRIES

Mainly the sales and financial management industry led the way for telecommuting, however, over the past few years much other business in U.S. have begun to implement telecommuting arrangements in several large cities. In Denver, Colorado, Bethesda, Maryland and Charlottesville, Virginia, IBM has furnished employees with furniture and equipment to work at home when they

are not in the field with customers. Recently, Ernst and young initiated in their Chicago accounting and consulting operation in 1992 “hoteling”. According to the UC-Davis study, in the first year Ernst & Young reduce space requirements by 7 percent, and the company expects to eventually shrink its office space requirements nationwide by two million square feet for saving of \$40 million a year.

Since the implementation of telecommuting can also bring benefit to the whole society, so it is necessary for

the government and local authorities to do something useful to support telecommuting.

In general, there are four primary forces that will benefit the public interest in promoting telecommuting: the environment, the transportation demand, economics, and quality of life. Some developed countries have made some policies to support telecommuting. (Examples: Table 2)

National governments in the European Union advocate telecommuting both as a means to reduce traffic congestion and increase women's labor force participation. Throughout the European Union, an average of 60% of employees shows an interest in it. The interest among those looking for work is even stronger. Employers giving their employees the opportunity to telecommute may be regarded as attractive.

In Singapore a study of 285 employees of an IT company confirmed that employees would be less likely to end the employment relationship if the company offered telecommuting as an option. Experimental research with students reaffirmed the claim that the concept of telecommuting will attract well-qualified individuals to join the organization.

In U.S., Equal Employment Opportunity Commission (EEOC) encourages work at home (telework) as a reasonable accommodation. If the disabled employee is able to perform the essential functions of the job at home, then telecommuting may be a required accommodation---unless the employer can find a way to accommodate the employee's disability otherwise or prove that telecommuting would cause undue hardship on the company.

3. SOME SUGGESTIONS BASED ON CHINA ACTUALITY

Based on the analysis of china actuality, there are some reasons for implementation of telecommuting and corresponding suggestions.

3.1 Attract elitist

China is a developing country and is developing very quickly. Especially during recent decades, the development of global information technology and the globalization accelerate china's speed to infiltrate the whole world. From the whole world, telecommuting makes it possible for the Chinese companies to attract more elites both at home and abroad, especially the ones oversea. The basic policy of government for supporting telecommuting is constructing, improving and perfecting the information speedway.

West development strategy is very important to the Chinese economic growth. In those regions, one of the big problems of economic development is lacking of

human resources, elites. Human resource is crucial to a country's economic growth. In contrast to cities in west and west middle regions, the developed cities in east regions following the line of the sea are full of all kinds of specialists. Although, there are a lot of persons of noble aspirations and young people full of passion have eagerly responded to the national call and devoted to the West and Middle West regions. Relative to the rich human resource in east regions, the number of persons who have devoted themselves to the development of West and Middle West is still very small. The reasons are various, but one is that because of less developed, the living standard in those regions is inferior to the East regions, especially the big cities of East regions following the line of the sea such as Shanghai, Shenzhen and so on, more people will choose to work or settle in developed regions in stead of West and Middle West regions, especially those whose hometown is far away the West and Middle West. How to attract those people work for the companies that site in the West and Middle West regions? The implementation of telecommuting comes up with the solution. For example, the company can provide the employees opportunities to work at their own places; these places may be the employees' home, the office in another city, or a telecenter.

In order to support telecommuting for such kind of reason, the government and local authorities can provide tax exemption or preferential duty. The beneficial tariff can be various. They can provide the beneficial tariff according to the ratio of the number of telecommuter versus the number of whole employees, also can according to the ratio of time which spent in work by telecommuting versus the whole time spent by the whole employees in the whole company work. The government and local authorities can also provide the telecommuters some kind of subsidy, which not only can encourage more people to work for those companies but also can lighten the financial burden of those companies.

3.2 Consideration for Female employees and disabilities

In China, more and more women accept high education and the percentage of female employees is very high.

Different from male, female have the responsibility to procreate and most women want to be a mother. Because of inconvenience of pregnancy and traffic jam, many pregnant women give up their job. Furthermore, many new mothers after giving birth to a baby still can not go back to work; sometimes they have to look for another new work, sometimes they have to take care of their new-birth baby. As far as the companies which the pregnant women have worked for concerned, on one hand they want to retain those experienced employees, on the other hand they can not accept their a long-time-leave, because sometime the off-work time will reach one year. Now, the Information Technologies (IT) and Information System (IS) provide opportunities

for organizations/companies to provide telecommuting to their employees who need such kind of work arrangement. So the telecommuting cannot only benefit for the company but also the pregnant employees and new mother employees. The move which would see changes in women's workload, employment schedule and the facilitation of "telework" at workplaces is to enable women's participation in the labor force and enable them to juggle between work and family responsibilities.

In view of situations above, the government can take some measures to encourage or to say, to some degree, oblige some companies to implement telecommuting. For example, when a female employee has been pregnant, and by carefully redesigning, her work can be done at home, if the employee applies for working at home, the employer cannot refuse. Once the employer fires the employee for this reason, the employer should be punished. This is a measure to protect women's right.

The Chinese government always pays attention to the disabilities' right and employment. The government encourages welfare organizations to centralized arrange disabilities' employment by tax exemption or preferential duty. At the same time, the government has set a certain hired disabilities ratio and required all the companies should employ a certain number disabilities according to the ratio. It is believed that by implementing telecommuting, more and more disabilities can find a job, especially for those who have difficulties in traveling.

With the increase of population in cities, communications has been becoming more and more crowded. Quite a number of employees have to spend two or more hours going to the office building from their home and back, not to mention when meeting the traffic jam. So, working parents may highly value the time-savings of telecommuting due to the elimination of commuting time, or because telecommuting may allow a parent to stay at home with a sick child.

As far as many medium-sized and small enterprises are concerned, on one hand a lot of tasks are suited to telecommuting, on the other hand, at the present, they have not the ability to spend a great deal of money to equip their offices and their employees' home offices. Furthermore, the most important is that they worry about the data security. Under such kind of condition, the government itself can build several telecenters or organize several big companies to set several telecenters. And any companies that need telecommuting arrangement can rent some offices in these telecenters.

Many people may have noticed that managers are more likely to accept a new program or innovation that has been widely and successfully used by other managers. So I think, in order to encourage all kinds of organizations to implementing telecommuting, the government department should be the first to set telecommuting arrangement and become an example in successful implementing telecommuting. For example, the

government can set a plan includes a goal to increase the percentage of government employees that telecommute to 6 percent by the end of some year say 2005. Another benefit of this measure is that once the telecommuting is successfully implemented in the government departments, which can provide experience and example the official who make the support policy and let him or her make more wise decision and more practical policy.

In spite of measures mentioned above, the government can also provide free training or financial support for the companies that are planning to implement telecommuting. For example, the government and local authorities may make budgetary allocations for installation of telephone lines, computers and related equipment and monthly funding the telecenters.

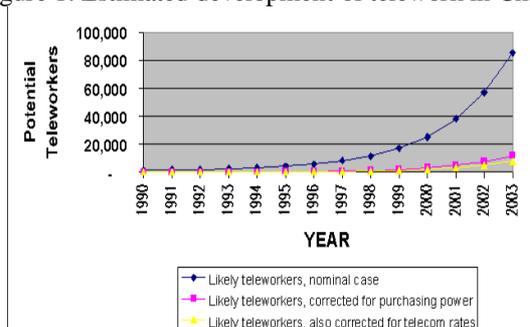
Certainly, the government and local authorities can provide various specific support policies (such as preferential telephone rate) for telecommuting implementation according to different conditions. Whether telecommuting in one company is suitable is akin to the employer, the employee and the task. For example, telecommuting is not suitable for employees who require intensive face-to-face interaction with coworkers and customers, or who require physical resources to handle their work.

In China, the environment to carry out telecommuting has not yet been well established. Although telecommuting is expected by the experts, for instance as a solution for the traffic congestion in the urban area, the concept of telecommuting has not spread over the peoples. So, from now on, not only the establishment of the telecommuting environment but also activities for recognition and enlightenment of telecommuting will be needed.

Several environmental factors are in place in China that would lead to successful implementation of telework programs. These include a proliferating telecommunications industry, more direct foreign investment and lifted trade restrictions that have created more jobs.

Although data regarding telework in China are very scarce, someone has made some predictions of the growth of telework, based on decades of experience in the US. The current forecast is shown in figure 1.

Figure 1: Estimated development of telework in China



Source: http://www.ecatt.com/country/china/inhalt_cn.htm

6. CONCLUSION

Implementing telework programs in China's public and private sectors could result in increased productivity with a declining workforce. Home-based telework programs could be the first solution for childcare and eldercare anxieties, and mobile telework can also play an important role.

In order to successfully implement telecommuting, the government must play an active role to help and encourage the companies to establish such kind of new work form.

Firstly, the government should help employer identify the jobs best suited to distance work. Some work lends itself naturally to off-site arrangements, such as sales, customer service, and auditing.

Secondly, the government should help employer approach telecommuting carefully by screening workers to determine eligibility and providing training in commonsense ways to avoid the potential risk of technological and financial failure of telecommuting program, decide which kinds of employees are suited to telecommuting, which kinds of employees are motivated, trustworthy and able to work independently.

Thirdly, the government should help employer develop guidelines, including approved reasons for telecommuting, using a personal vehicle for company business, and so on.

Finally, the government should help employer provide his or her employees an appropriate telework office with higher-speed broadband, cost-effective video that in a space totally dedicated to the work at hand and designed for comfort and injury prevention.

ACKNOWLEDGEMENT

The paper is supported by NSFC Project 70271025 and

70121001.

REFERENCES

- [1]. Pascale Peters, Kea G. Tijdens, Cecile Wetzels. Employees' opportunities, preferences, and practices in telecommuting adoption. *Information & Management* Volume: 41, Issue: 4, March 2004, pp. 469-482.
- [2]. Mills, Juline E.; Wong-Ellison, Chilian; Werner, William; Clay, Joan M. Employer Liability for Telecommuting Employees, *Cornell Hotel & Restaurant Administration Quarterly*, Oct2001, Vol. 42 Issue 5, p48, 12p, 4bw.
- [3]. Disability Law, Fair Employment Practices Guidelines, 4/1/2003 Issue 573, p5, 2pEEOC.
- [4]. Kistner, Toni, Big ideas for advancing telework, *Network World*, 11/10/2003, Vol. 20 Issue 45, p39, 5/6p, 1c.
- [5]. Christos J. Georgiou, Petros S. Stefanias, Strategies for Accelerating the Worldwide Adoption of E-Commerce, *COMMUNICATIONS OF THE ACM*, April 2002/vol.45.No.4ve, pp.145-151.
- [6]. Will flexible working policies confirm that a woman's place is in the home? , *Management Services* December 2003, pp: 20-23
- [7]. Kistner, Toni. Take control of Telework Support, *Network World*, 3/17/2003, Vol. 20 Issue 11, p31, 2/3p.
- [8]. Vivien K. G. Lim; Thompson S. H. Teo. To work or not to work at home-An empirical investigation of factors affecting attitudes towards telework. *Journal of Managerial Psychology*, Volume: 15 No. 6; 2000.
- [9]. Toshiya Jitsuzumi; Hitoshi Mitomo; Hajime Oniki. ICTs and sustainability: The managerial and environmental impact in Japan. *Foresight*, Volume: 3 No.2; 2001.
- [10]. Kunihiko Higa; Bongsik Shin, The telework Experience in Japan. *Communications of the ACM* (September 2003) Volume: 46 Issue: 9.
- [11]. Ursery, Stephen. Austin fights air pollution with telework program. *American City & County*, 0149337X, May2003, Vol. 118, Issue 5.
- [12]. GSA Offering Free Trial Runs at Its Telework Centers. *Washington Post*, the, 01908286, Jan 26, 2004.
- [13]. State telecommuting center offers free test period to federal agencies, *Daily Record*, The (Baltimore, MD), Feb 04, 2004.
- [14]. Manuela Perez Perez; Angel Martinez Sanchez; Maria Pilar de Luis Carnicer, The organizational implications of human resources managers' perception of telework. *Personal Review* Volume: 32 No.6; 2003
- [15]. Nancy B. Kurland, Diane E. Bailey; telework: the advantages and challenges of working here, there, anywhere, and anytime, *Organizational dynamics*, autumn 1999, p53-68

The New Vigilantism: Combating the Piracy of Copyrighted Materials on Peer-to-Peer Networks

Alvin Stauber

College of Business, Florida State University, Tallahassee, Florida 32306-1110, U.S.A.
astaube@cob.fsu.edu

ABSTRACT

Non-existent just a few years ago, peer-to-peer networks have experienced phenomenal growth. One consequence of the expansion of these networks has been extensive unauthorized downloads of copyrighted materials. Losses attributable to music piracy alone are estimated to exceed \$4 billion. In response to this problem and to combat the misappropriation of copyrighted materials, the Peer-To-Peer Piracy Prevention Act ("Act") was introduced in the United States Congress in 2002. The aim of the law is to enable copyright holders to utilize technological self-help measures to stop copyright infringement on peer-to-peer networks. This paper describes the developments leading to the introduction of the Act and briefly analyzes the provisions of the Act. In addition, the paper discusses the legal issues raised by the self-help remedies contained in the Act and concludes that there is ample precedent to support the legal propriety of these measures.

Keywords: vigilantism, copyright, peer-to-peer network, self-help remedies, citizen's arrest, safe harbor

1. INTRODUCTION

"Hollywood Vigilantes v. Copyright Pirates"
(Humphrey, 2003)

"Streamcast Opposes . . . Posse of Copyright
Vigilantes" (Streamcast, 2002)

"Vigilantes v. Pirates: The Rumble Over Peer To Peer
Technology" (Fazekas, 2003)

The foregoing headlines provide a glimpse into the battle that is being waged over the misappropriation of copyrighted materials on peer-to-peer networks. Words such as "vigilantism" and "piracy"—ordinarily reserved for portraying mischief in the Wild West or on the wild seas—are now part of the lexicon used to describe this cyber warfare. One of the principal contributors to this adversarial atmosphere was the proposed enactment in the United States of legislation known as the Peer-To-Peer Piracy Prevention Act (H.R. 5211, 2002). As one observer stated, it was "greeted by a swirl of controversy in the Internet community" (Humphrey, 2002). The purpose of this paper is: (1) to describe the developments that led to the introduction in the United States Congress of the Peer To Peer Piracy Prevention Act ("Act"); (2) to briefly analyze the provisions of the Act; and (3) to discuss the legal issues raised by the self-help remedies contained in the Act.

2. DEVELOPMENTS

Non-existent just a few years ago, peer-to-peer networks have experienced phenomenal growth. The leading peer-to-peer file sharing software has been downloaded over 200 million times, and its developer claims over 60 million users (Davidson, 2003). Peer-to-peer file sharing networks are different from other Internet applications in the following respects: they

tend to share data from a large number of end user computers rather than from the more central computers generally thought of as Web servers. A key innovation of peer-to-peer file sharing networks is their sophisticated mechanisms for searching millions of "shared" files to find data among many connected systems. Information on peer-to-peer networks tends to be less centrally controlled and more reflective of what end user participants believe is valuable or worth sharing (Davidson, 2003).

Napster first brought peer-to-peer networks into the limelight, but it was shut down by the courts because it facilitated—using a central directory and centralized servers—massive copyright infringement. Current peer-to-peer networks avoid the "Napster problem" by incorporating varying levels of decentralization. While unauthorized downloads of copyrighted software, games, photographs, tapes, and movies have occurred, the most rampant piracy of copyrighted work involves music. Losses attributable to music piracy are estimated to amount to \$4.3 billion annually (Sorkin, 2003).

Most of the efforts to thwart music piracy have focused on litigation instituted against companies and individuals that are involved in the distribution of pirated music. Last year, four college students settled a lawsuit initiated by the music industry by agreeing to halt their music swapping network and agreeing to pay penalties of \$12,000 or more. And recently a federal judge ruled that Verizon Communications must disclose to the Recording Industry Association of America the identities of customers suspected of distributing pirated music (Sorkin, 2003). But copyrighted works are spread over peer-to-peer networks so quickly that they wind up almost instantly on millions of computers, rendering litigation against

individual peer-to-peer users essentially ineffective. It is in response to these developments that the Peer To Peer Piracy Prevention Act was introduced by California Congressman Howard Berman in the United States Congress on July 25, 2002.

3. PEER TO PEER PIRACY PREVENTION ACT

In introducing the proposed legislation, Representative Berman explained the thrust of the Act as follows:

[E]nactment of the legislation I introduce today is necessary to enable responsible usage of technological self-help measures to stop copyright infringements on peer-to-peer networks . . . There is nothing revolutionary about property owners using self-help—technological or otherwise—to secure or repossess their property. Satellite companies periodically use electronic countermeasures to stop the theft of their signals and programming. Car dealers repossess cars when the payments go unpaid. Software companies employ a variety of technologies to make software non-functional if license terms are violated (Berman, H.L., 2002).

Lamenting the fact that “the primary current application of peer-to-peer networks is unbridled copyright piracy,” Representative Berman sought in the Act to permit copyright owners to employ a variety of technological tools—e.g., interdiction, decoys, redirection, file-blocking, and spoofs—to prevent the illegal distribution of copyrighted works over a peer-to-peer network (Berman, H.L., 2002). The protection afforded by the Act is to provide a “safe harbor” from liability when copyright owners employ self-help measures to prevent piracy of their works. But these self-help measures are narrowly tailored to ensure “acceptable behavior” by the copyright owner. Self-help measures that would involve planting a virus on a peer-to-peer user’s computer or otherwise removing, corrupting, or altering files or data on the user’s computer are prohibited and are subject to civil and administrative remedies (H.R. 5211, 2002).

In addition, the copyright owner is denied the safe harbor from liability if:

- (1) The copyright owner impairs the trading of files that do not contain any portion of the copyrighted work, unless such impairment is “reasonably necessary” to impair the trading of the copyrighted work;
- (2) The actions of the copyright owner cause economic loss to anyone other than the file trader;
- (3) The actions undertaken by the copyright owner cause more than fifty dollars of economic loss to the file trader, other than loss involving the copyrighted works; or
- (4) The copyright owner does not provide the required 7-day notice to the Department of Justice disclosing the

proposed method to be used to stop the copyright infringement (H.R. 5211, 2002; Humphrey, 2003).

4. LEGAL ISSUES

Immediately after its introduction, the Act was attacked by critics who claimed that the self-help remedies contained in the proposed legislation might result in a type of vigilantism in which copyright owners—taking the law into their own hands—might “hack” into computers, damaging the hardware, compromising legitimately downloaded files, and altering data. In addition, critics raised constitutional concerns, claiming that the self-help measures would violate the privacy of individual users (Streamcast, 2002).

4.1 Vigilantism

To characterize the self-help measures proposed by Congressman Berman as “vigilantism” is hyperbole of the first order. Society undoubtedly welcomes the prospect of a “community of ‘vigilant’ citizens who watchfully and lawfully seek to detect and prevent crime” (Brandon, et al., 1984). Vigilantism, on the other hand, is defined as an “organized, extra-legal movement, the members of which take the law into their own hands” (Sklansky, 1999). Vigilantes typically engage in the violent exercise of police power authority in an unlawful manner; they “violate the law, often heinously, in the name of law and order” (Brandon, et al., 1984). Being vigilant, however, does not make one a vigilante. The Peer-To-Peer Piracy Prevention Act emphatically rejects any measures that are violent or heinous or unlawful. Indeed, odious behavior such as planting a virus, altering or corrupting files, or “freezing” a computer system is proscribed under the proposed law and would result in civil and administrative remedies (H.R. 5211, 2002). What the Act advocates are self-help measures that do not disrupt the technical operation of a person’s computer or networks. The Act encourages vigilance and self-help, not vigilantism.

If anything, the self-help remedies envisioned by the Act are the antithesis of the vigilantism of bygone years. The vigilante activity that first appeared in the United States in the late 1700s involved “the frontier justice afforded cattle rustlers, horse thieves, murderers, thugs, and desperados” (Brandon, et al., 1984). Students of vigilantism, while acknowledging that our legal system condemns the practice, nevertheless warn that:

society should heed the message that outbreaks of this behavior sometimes suggest. Although many instances of vigilantism probably reflect extremist behavior, some vigilante activity also may suggest a latent societal feeling of dissatisfaction with the operation of criminal laws and the justice system (Brandon, et al., 1984).

It is, indeed, the dissatisfaction with the operation of our legal system that motivated Congressman Berman and others to try to improve that system by seeking legal approbation of self-help measures employed by copyright owners for their own protection. If this is vigilantism, it is a 21st century version and ought to be re-named the “new vigilantism”, in recognition that it is a practice that advocates self-help through legal, as opposed to illegal, means, and is, therefore, significantly different from the “old vigilantism”.

4.2 Self-help Remedies

Traditionally, self-help remedies have been sanctioned by the law. As indicated earlier, satellite companies use electronic countermeasures to stop theft of their signals and programming; retailers repossess goods when loans go into default; and software companies employ a variety of technologies to make software non-functional if license terms are violated. In the case of repossession of goods, the retrieval of merchandise does not have to be carried out by judicial order or by law enforcement officers' action. Rather, upon default, the creditor or his agent is authorized to enter private or public property so long as the taking of the item to be retrieved is effected without a “breach of the peace”. Entering the debtor's private property is not usually considered a trespass (thereby eliminating the possibility of a breach of the peace), because although a lack of consent to the entry might be claimed, there is nevertheless a privilege to enter—rendering consent unnecessary—created by the repossession right (Brandt v. Dugent, 2000). In fact, in the case of automobile repossession, the retrieving person is authorized to “break into” the vehicle in order to effect the repossession, although liability for damage to the car might be incurred (Clarke, 2001; Foster, 1982). This lawful procedure involving self-help by entering the real and personal property of another provides a precedent for the self-help measures in the Act.

Another precedent that provides support for the Act's remedies is that found in the law of citizen's arrest. The self-help power of citizen's arrest is available in most states to those private persons and businesses that observe a theft or other crime being committed. In fact, the laws of at least six states in the United States permit an individual to use force or to break a door or window to effect a citizen's arrest and thereby possibly effect recovery of stolen property (Code of Ala., 2004; A.R.S., 2004; HRS, 2003; Miss. Code Ann., 2004; N.D. Cent. Code, 2003; Tenn. Code Ann., 2004). The concept of “breaking into” another's property has relevance to the self-help measures advocated by the Act. Indeed, one commentator described the Act as “a bill that would have given the film and music industries immunity from prosecution if they found ways to **break into the swappers' computers** (emphasis supplied) and block their swapping” (Farrar, 2003). Admittedly, breaking into an automobile to

recover tangible property differs from breaking into a computer to recover intangible (intellectual) property, but the underlying concept remains the same. Critics of the Act's self-help measures argue that the differences are significant, stating as follows:

There are important differences between the two given the more utilitarian construction of intellectual property rights in the Constitution. IP rights are more limited in scope and duration, and the “fair use” rights of users is part of today's copyright bargain. Then again, fair use doesn't necessarily mean free use. Artists deserve compensation (Farrar, 2003).

Predictably, Representative Berman rejects the notion that “breaking and entering” into a person's computer is sanctioned by the Act. He explains his position as follows:

Despite wildly inaccurate press reports, H.R. 5211 in no way allows a copyright owner to “hack” into anyone's computer. Copyright owners are only allowed to enter or look into a P2P user's computer to the same extent that any other P2P user is able to do so. In other words, if a KaZaA user has advertised to all 100 million other KaZaA users that he wants to download or distribute a copyrighted song, the songwriter is not “hacking” if she reads the advertisement like everyone else. H.R. 5211 then allows the songwriter to take certain, limited actions to stop the distribution of her copyrighted song between KaZaA users, but in no way allows her to enter or look into a private area of those KaZaA users' computers (Berman, H., 2002).

Another complaint leveled by critics is that self-help copyright holders can make mistakes when they take action against alleged infringers (Davidson, 2003). Interestingly enough, the analogy to the self-help remedy of citizen's arrest again provides useful insight into how this problem is dealt with. According to most state statutes, a retail merchant has the right to detain a suspect (i.e., effect a citizen's arrest) whom the merchant reasonably believes has engaged in shoplifting as long as that detention is done in a reasonable manner (Ohlin & Stauber, 2003). Illustrative is the Florida law, which provides as follows:

a merchant . . . who has probable cause to believe that a retail theft . . . has been committed by a person and . . . that the property can be recovered by taking the offender into custody may, for the purpose of attempting to effect such recovery or for prosecution, take the offender into custody and detain the offender in a reasonable manner for a reasonable length of time.

The taking into custody and detention by a . . . merchant, merchant's employee, or . . . agent, if done in compliance with all the requirements of this

subsection, shall not render such [person] criminally or civilly liable for false arrest, false imprisonment, or unlawful detention (Fla. Stat. Sec. 812.015(3), 2004).

In other words, the Florida statute provides a “safe harbor” for merchants who comply with the requirements of the law. That is, criminal and civil immunity is granted to merchants so long as they act **reasonably**, even if they mistakenly accuse someone of shoplifting. While erroneously accusing a person of retail theft can understandably result in a terribly unfortunate experience for the innocent suspect, the law providing merchants with immunity balances the right of merchants who undertake reasonable measures to apprehend suspected thieves with the rights of an innocent person who is mistakenly accused of theft. Because shoplifting has become a problem of epidemic proportions, the law tips the scales in favor of the merchant (Sparrow, 2003).

A similar argument can be made in the situation involving on-line theft of copyrighted materials. So long as the copyright owner’s impairment of the unauthorized distribution or reproduction of the copyrighted work does not “alter, delete, or otherwise impair the integrity of any computer file or data residing on the computer of a file trader”, the copyright owner incurs no liability (H.R. 5211, 2002). Similar to the reasonableness criterion of the merchant shoplifter apprehension statutes, the Act in Section 514(b) creates a safe harbor exception. According to the Act, impairing the trading of files that do not contain any portion of the copyright owner’s copyrighted work would result in liability unless such impairment is “reasonably necessary” (H.R. 5211, 2002). Critics of the proposed law contend that such language creates serious problems. As one critic observed: the ambiguity of the words “reasonably necessary” paves the road to future litigation over the issue. . . . Although [Representative Berman] has referred to the “reasonably necessary” exception as “certain necessary circumstances,” [he] has yet to reveal what those circumstances are. Considering that H.R. 5211 explicitly allows copyright owners to navigate around existing state and federal law, Representative Berman’s desire to create legislation that is “narrowly crafted, with strict bounds on acceptable behavior by the copyright owner” will not be achieved until this provision is made more clear and the ability of file traders to trade legally is adequately preserved (Humphrey, 2003).

It should be recognized, however, that many statutes—including the merchant shoplifter apprehension statutes earlier described—utilize a “reasonableness” standard to guide individuals in their conduct. Indeed, there are entire bodies of law—most notably the law of negligence in its utilization of the reasonable person standard—that are principally based on the concept of reasonableness. The term “reasonable” is used in

thousands of contexts in the law; its use simply acknowledges that all fact situations are different and must be viewed in their own particular contexts. The definition of what is reasonable in any given set of circumstances is, therefore, situation-specific. Incorporating a reasonableness standard into a law is not fatal to the law. It simply means that the finder of fact—a judge or jury—will have to make the determination as to whether a particular course of conduct is reasonable under the circumstances (Clarkson, et al., 2004).

With regard to constitutional concerns relating to the Act’s self-help measures and their potential for invasion of privacy, the focus appears to be on the Fourth Amendment’s prohibition of illegal searches and seizures. The problem with attacking the Act’s self-help measures on this basis is that, as many courts have stated, “The Fourth Amendment’s proscription against warrantless searches and seizures does not apply to searches by private individuals not acting as agents of the state” (State v. Cooney, 1995). The copyright owners seeking to protect themselves against the piracy of their copyrighted work through self-help measures thus appear to have some latitude that law enforcement officers and/or government officials would not necessarily enjoy.

5. CONCLUSION

The self-help measures embodied in the Act are innovative and potentially effective weapons in the battle to combat the piracy of copyrighted work. While certainly subject to abuse, these measures are restricted in the Act so as not to result in harm to the hardware or software of peer-to-peer users. Although these self-help remedies are subject to constitutional and other legal concerns, there is ample precedent to support the legal propriety of these measures.

REFERENCES

- [1] A.R.S., Sec. 13-3892, 2004.
- [2] Berman, H., “Frequently asked questions about P2P Piracy Prevention Act,” at http://www.house.gov/berman.p2p_faq.html, July 25, 2002.
- [3] Berman, H.L., “Introduction of the Peer to Peer Piracy Prevention Act,” 107th Congress, at http://www.house.gov/berman/newsroom/piracy_prevention_act.html, July 25, 2002.
- [4] Brandon, D.I., et al., “Self-help: extrajudicial rights, privileges and remedies in contemporary american society,” *Vanderbilt Law Review*, Vol. 37, pp. 845-937, May 1984.
- [5] Brandt v. Dugent Liquidating Trust, 749 So.2d 1276, 2000 Fla. App. LEXIS 1675, 2000.
- [6] Clarke, S.S., “Repossession soars as economy plunges,” *Tallahassee Democrat*, p. 4B, December 31, 2001.
- [7] Clarkson, K.W., et al., *West’s Business Law*, 9th ed., Mason, Ohio: Thomson, 2004.
- [8] Code of Ala., Sec. 15-10-7, 2004.
- [9] Davidson, A., “Peer-to-peer file sharing privacy and security,” Testimony before U.S. House Committee on

Government Reform, at <http://www.cdt.org/testimony/030515davidson.shtml>, May 15, 2003.

[10] Farrar, C., "Hatch backs off P2P hack approval . . . somewhat," at http://www.avn.com/index.php?Primary_Navigation=Articles&Action=View_Article&Co, June 18, 2003.

[11] Fazekas, C., "Vigilantes v. pirates: the rumble over peer-to-peer technology hits the house floor," *Duke Law and Technology Review*, Vol. 2002, p. 20.

[12] Fla. Stat., Sec. 812.015(3), 2004.

[13] Foster, J.E., "Parties' remedies on default," *Secured Transactions in Florida*, 1982.

[14] H.R. 5211, 107th Congress, 2002.

[15] HRS, Sec. 803-11, 2003.

[16] Humphrey, J.S., "Debating the proposed Peer-to-Peer Piracy Prevention Act: should copyright owners be permitted to disrupt illegal file trading over peer-to-peer networks?" *North Carolina Journal of Law and Technology*, Vol. 4, pp. 375-417, 2003.

[17] Miss. Code Ann., Sec. 99-3-11, 2004.

[18] N.D. Cent. Code, Sec. 29-06-22, 2003.

[19] Ohlin, J.B., Stauber, A., "The applicability of citizen's arrest powers to the hospitality industry," *Journal of Hospitality & Tourism Research*, Vol. 27, No. 3, pp. 342-357, 2003.

[20] Robinson, T., "Bill would let copyright owners infiltrate file-swapping services," *E-Commerce Times*, July 26, 2002, at <http://www.ecommercetimes.com/story/18766.html>.

[21] Sklansky, D.A., "The private police," *UCLA Law Review*, Vol. 46, pp. 1165-1286, April 1999.

[22] Sorkin, A., "Software bullet is sought to kill music piracy," *N.Y. Times*, p. 1, May 4, 2003.

[23] Sparrow, A., "Shoplifting has become epidemic," *The Daily Telegraph*, July 31, 2003.

[24] State v. Cooney, 320 S.C. 107, 463 S.E. 2d 597, 1995.

[25] "Streamcast networks opposes Congressman Howard Berman's proposal calling for posse of copyright vigilantes," at http://www.streamcastnetworks.com/6_25_02.html, June 25, 2002.

[26] Tenn. Code Ann., Sec. 40-7-112, 2004.

The Security Flaw of an Untraceable Signature Scheme

Chinchen Chang, Yafen Chang

Department of Computer Science and Information Engineering, National Chung Cheng University,
Chiayi 621, Taiwan, China
e-mail: {ccc, cyf}@cs.ccu.edu.tw

ABSTRACT

In 2003, Hwang et al. proposed a new blind signature based on the RSA cryptosystem. The Extended Euclidean algorithm is employed in their proposed scheme. They claimed that the proposed scheme was untraceable and it could meet all requirements of a blind signature. However, we find that the signer can still trace the blind signature applicant in some cases. Thus, we present the security flaw of Hwang et al.'s scheme in this paper.

Keywords: blind signature, untraceable, RSA, Extended Euclidean algorithm

1. INTRODUCTION

In 1982, Chaum proposed the concept of blind signature [1]. In blind signature schemes, an applicant can obtain a signature of a message from the signer without revealing the content of the signed message to the signer. Blind signature can be used in many cryptographic applications, such as electronic voting systems and electronic cash payment systems. Thus, how to make the resulting message-signature pair not be able to be linked is an important issue. On the other hand, the personal information should be protected while the resulting message-signature pair is used in any application. As a result, Chaum proposed the first blind signature ensuring the user's private information. With the progressive improvement of the blind signature [2], [4]-[6], the requirements of the blind signature, (1) correctness, (2) blindness, (3) unforgeability, and (4) untraceability, are listed and described as follows:

- (1) Correctness: Anyone can check the blind signature of the signed message by using the server's public key.
- (2) Blindness: The signer has no idea of the content of the signed message.
- (3) Unforgeability: Only the signer can generate the signature. That is, no one can forge a valid signature and can have the forged signature verified successfully.
- (4) Untraceability: The signer of the blind signature cannot link the message-signature pair even if the signature has been revealed to be public.

Recently, Hwang et al. [7] proposed a blind signature based on the RSA cryptosystem [9]. It also employs the Extended Euclidean algorithm [8]. They claimed that the proposed scheme is untraceable and meets all requirements of a blind signature mentioned above. And the security of the proposed scheme is based on the difficulties of solving the factoring problem. However, in some cases, the signer can still trace the blind signature applicant. As a result, we present its security flaw in this paper.

The rest of the paper is organized as follows. First, we

review Hwang et al.'s untraceable blind signature in Section 2. Then the drawback of Hwang et al.'s scheme is shown in Section 3. Finally, the conclusions are given in Section 4.

2. A REVIEW OF HWANG ET AL.'S UNTRACEABLE BLIND SIGNATURE

In this section, we review Hwang et al.'s proposed untraceable blind signature, which consists of five phases: (1) the initialization phase, (2) the blinding phase, (3) the signing phase, (4) the unblinding phase, and (5) the verification phase. The five phases are presented in Subsections 2.1 to 2.5, respectively.

2.1 The Initialization Phase

In this phase, the signer S makes the essential information public as follows:

- Step 1. S randomly selects two large prime numbers p and q and computes $n = p \cdot q$ and $\phi(n) = (p-1)(q-1)$.
- Step 2. S chooses two large random numbers e and d , where $\gcd(e, \phi(n)) = 1$, $e \cdot d \bmod \phi(n) = 1$.
- Step 3. S keeps p , q , and d secret and makes e , n , and H public, where H is a one-way hash function—MD5 and SHA-1 [3] for example.

2.2 The Blinding Phase

Suppose the requester R has a message m and wants m signed without being known by S . R performs the following steps to make m concealed.

- Step 1. R randomly selects two distinct numbers t_1 and t_2 .
- Step 2. R chooses two random primes a_1 and a_2 , where $\gcd(a_1, a_2) = 1$.
- Step 3. R computes $s_1 = t_1^e \cdot H(m)^{a_1} \bmod n$ and $s_2 = t_2^e \cdot H(m)^{a_2} \bmod n$.
- Step 4. R sends s_1 and s_2 to S .

2.3 The Signing Phase

After getting s_1 and s_2 from R, S generates the corresponding blind signature of m as follows:

Step 1. S randomly chooses two primes b_1 and b_2 , where $\gcd(b_1, b_2) = 1$.

Step 2. S computes $r_1 = s_1^{b_1 d} \text{ mod } n$ and $r_2 = s_2^{b_2 d} \text{ mod } n$.

Step 3. S sends (r_1, r_2, b_1, b_2) to R.

2.4 The Unblinding Phase

Upon receiving (r_1, r_2, b_1, b_2) , R performs as follows to derive the blind signature s of m.

Step 1. R computes $g_1 = r_1 \cdot t_1^{-b_1} \text{ mod } n$ and $g_2 = r_2 \cdot t_2^{-b_2} \text{ mod } n$.

Step 2. R finds w and t according to Extended Euclidean Algorithm [3], where $(a_1 b_1)w + (a_2 b_2)t = 1$, and keeps b_1, b_2, w , and t secret.

Step 3. R computes $s = g_1^w \cdot g_2^t \text{ mod } n$ and then publishes (m, s) .

2.5 The Verification Phase

In order to verify the signature s of m, the verifier V computes $H(m)$ and $s^e \text{ mod } n$. Then V checks whether $H(m) = s^e \text{ mod } n$ holds or not. If it holds, it denotes that s is indeed the signature of m.

3. THE SECURITY FLAW OF HWANG ET AL.'S UNTRACEABLE BLIND SIGNATURE

In this section, we are going to show how the signer traces the blind signature in Hwang et al.'s proposed scheme. To make tracing the blind signature easier, S chooses two primes p and q, where $4|p+1$ and $4|q+1$, and computes $n = p \cdot q$ and $\phi(n) = (p-1)(q-1)$. Then S randomly chooses two large numbers e and d, where $\gcd(e, \phi(n)) = 1, e \cdot d \text{ mod } \phi(n) = 1$.

First, as shown in the blinding phase, R has a message m and wants m signed without being known by S. Then, R performs as follows:

Step 1. R randomly selects two distinct numbers t_1 and t_2 .

Step 2. R chooses two random primes a_1 and a_2 , where $\gcd(a_1, a_2) = 1$.

Step 3. R computes $s_1 = t_1^e \cdot H(m)^{a_1} \text{ mod } n$ and $s_2 = t_2^e \cdot H(m)^{a_2} \text{ mod } n$.

Step 4. R sends s_1 and s_2 to S.

Second, as shown in the signing phase, S generates the corresponding blind signature of m as follows:

Step 1. S chooses two random primes b_1 and b_2 , where $\gcd(b_1, b_2) = 1$.

Step 2. S computes $r_1 = s_1^{b_1 d} \text{ mod } n$ and $r_2 = s_2^{b_2 d} \text{ mod } n$.

Step 3. S sends (r_1, r_2, b_1, b_2) to R.

Third, as shown in the unblinding phase, R gets (m, s) , where $s = H(m)^d \text{ mod } n$. After performing the above procedures several times, S can get (s_1, s_2) 's and $(s_1^d \text{ mod } n, s_2^d \text{ mod } n)$'s. Because $s_1 = t_1^e \cdot H(m)^{a_1} \text{ mod } n$ and $s_2 = t_2^e \cdot H(m)^{a_2} \text{ mod } n$, we can have $s_1^d = t_1^{e \cdot d} \cdot (H(m)^d)^{a_1} \text{ mod } n$ and $s_2^d = t_2^{e \cdot d} \cdot (H(m)^d)^{a_2} \text{ mod } n$. That is, S can collect all the $(t_1^{e \cdot d} \cdot (H(m)^d)^{a_1} \text{ mod } n, t_2^{e \cdot d} \cdot (H(m)^d)^{a_2} \text{ mod } n)$'s.

Now, suppose that S knows (m', δ) , where $\delta = H(m')^d \text{ mod } n$. If t_1, t_2 , and $(H(m)^d \text{ mod } n)$ are co-prime and $a_1 < a_2$ possibly, S can find the relation between $(s_1^d \text{ mod } n, s_2^d \text{ mod } n)$ and δ as follows:

Step 1. S computes $\gcd(t_1^{e \cdot d} \cdot (H(m)^d)^{a_1} \text{ mod } n, t_2^{e \cdot d} \cdot (H(m)^d)^{a_2} \text{ mod } n) = H(m)^{d \cdot a_1} \text{ mod } n$.

Step 2. S computes $\eta = (H(m)^{d \cdot a_1} \text{ mod } n) \cdot \delta \text{ mod } n$.

Step 3. S computes

$$c_1 = \eta^{(p+1)/4} \text{ mod } p,$$

$$c_2 = (p - \eta^{(p+1)/4}) \text{ mod } p,$$

$$c_3 = \eta^{(q+1)/4} \text{ mod } q,$$

$$c_4 = (q - \eta^{(q+1)/4}) \text{ mod } q,$$

$$x = q(q^{-1} \text{ mod } p), \quad y = p(p^{-1} \text{ mod } q),$$

$$\beta_1 = (xc_1 + yc_3) \text{ mod } n,$$

$$\beta_2 = (xc_1 + yc_4) \text{ mod } n,$$

$$\beta_3 = (xc_2 + yc_3) \text{ mod } n, \text{ and}$$

$$\beta_4 = (xc_2 + yc_4) \text{ mod } n \quad [8].$$

Step 4. If there exists a β_j such that $\beta_i \cdot \delta^{(\phi(n)/2)} = \beta_j \text{ mod } n$, where $i \neq j$ and $1 \leq i, j \leq 4$, this denotes that δ is related to $(t_1^{e \cdot d} \cdot (H(m)^d)^{a_1} \text{ mod } n, t_2^{e \cdot d} \cdot (H(m)^d)^{a_2} \text{ mod } n)$.

If $m = m'$, we have

$$\eta = (H(m)^d)^{a_1 + 1} \text{ mod } n. \tag{1}$$

Because a_1 is odd, $a_1 + 1$ is an even number. As a result,

$$\eta = ((H(m)^d)^{(a_1 + 1)/2})^2 \text{ mod } n. \tag{2}$$

Equation (2) can be rewritten as the follow equation.

$$\eta = (((H(m)^d)^{(a_1 + 1)/2})^2 \text{ mod } n) \cdot (H(m)^{\phi(n)} \text{ mod } n) \text{ mod } n. \tag{3}$$

Since $m = m'$, we have

$$\eta = (((H(m)^d)^{(a_1 + 1)/2})^2 \text{ mod } n) \cdot (H(m')^{\phi(n)} \text{ mod } n) \text{ mod } n. \tag{4}$$

$$= (((H(m)^d)^{(a_1 + 1)/2} \text{ mod } n) \cdot (H(m')^{\phi(n)/2} \text{ mod } n))^2 \text{ mod } n. \tag{5}$$

According to the above equation, we can get

$$\eta^{1/2} = ((H(m)^d)^{(a_1 + 1)/2} \text{ mod } n) \cdot (H(m')^{\phi(n)/2} \text{ mod } n) \text{ mod } n. \tag{6}$$

From Equation (1), we have

$$\eta^{1/2} = ((H(m)^d)^{(a_1 + 1)/2} \text{ mod } n). \tag{7}$$

According to the properties of Rabin's [8], we know

there exist at most four distinct solutions for $\eta^{1/2} \pmod n$. That is, at least one β_i will equal to $((H(m)^d)^{(a_1^{-4})/2} \pmod n)$ for $1 \leq i \leq 4$. Therefore, if $m = m'$, we have

$$\beta_j = \beta_i * (H(m')^{\phi(n)/2} \pmod n) \pmod n \quad (8)$$

$$= \beta_i * \delta^{\phi(n)/2} \pmod n. \quad (9)$$

As a result, S checks whether any $\beta_i * \delta^{\phi(n)/2} = \beta_j \pmod n$ for $1 \leq i, j \leq 4$ and $i \neq j$, in Step 4.

According to the above procedures, it is obvious that S can trace the blind signature in Hwang et al.'s proposed blind signature scheme.

4. CONCLUSIONS

Hwang et al. proposed a new blind signature based on the RSA cryptosystem by employing the Extended Euclidean algorithm. Though they claimed that the proposed scheme was untraceable and it could meet all requirements of a blind signature, however, we find that the signer can still trace the blind signature applicant in some cases. On the other hand, the computation load of Hwang et al.'s scheme is too heavy. There still exists space for improving the proposed blind signature scheme.

REFERENCES

[1] Chaum, D., "Blind Signatures for Untraceable Payments," Proceedings of Advances in Cryptology

Crypto'82, Santa Barbara, California, USA, pp. 199-203, 1982.

[2] Chaum, D., "Blinding Signatures System," Proceedings of Advances in Cryptology Crypto'83, Santa Barbara, California, USA, pp. 153-156, 1983.

[3] Menezes, A. J., van Oorschot, P. C., and Vanstone, S. A., Handbook of Applied Cryptography, CRC Press, 1996.

[4] Fan, C.-I. Chen, W.-K., and Yeh, Y.-S., "Randomization Enhanced Chaum's Blind Signature Scheme," Commuter Communications, Vol. 23, pp. 1677-1680, 2000.

[5] Juang, W.-S. and Lei, C.-L., "Partially Blind Threshold Signatures Based on Discrete Logarithm," Computer Communications, Vol. 22, pp. 73-86, January 1999.

[6] Shao, Z., "Improved User Efficient Blind Signature," Electronics Letters, Vol. 36, No. 16, pp. 1372-1374, 2000.

[7] Hwang, M.-S., Lee, C.-C., and Lai, Y.-C., "An Untraceable Blind Signature Scheme," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E86-A, No. 7, pp. 1902-1906, July 2003.

[8] D. Kahn, The Codebreakers: The Story of Secret Writing, Macmillan Publishing Co., New York, 1967.

[9] Chang, C.-C. and Hwang, M.-S., "Parallel Computation of the Generating Keys for RSA Cryptosystems," IEE Electronics Letters, Vol. 32, No. 15, pp. 1365-1366, 1996.